Climate Data and Monitoring
WCDMP-No. 71

Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases

(Budapest, Hungary, 29 May – 2 June 2006)

World Meteorological Organization
Weather • Climate • Water
WMO-TD No. 1493

PROCEEDINGS OF THE FIFTH SEMINAR
FOR HOMOGENIZATION AND QUALITY
CONTROL IN CLIMATOLOGICAL DATABASES

Budapest, Hungary, 29 May – 2 June 2006

Organized by Hungarian Meteorological Service (HMS)
Supported by WMO and HMS

Edited by M. Lakatos, T. Szentimrey, Z. Bihari and S. Szalai

Chairperson, Publications Board
World Meteorological Organization (WMO)
7 bis, avenue de la Paix
P.O. Box No. 2300
CH-1211 Geneva 2, Switzerland

Tel.: +41 (0)22 730 84 03
Fax: +41 (0)22 730 80 40
E-mail: Publications@wmo.int

PREFACE

The First Seminar on Homogenisation was organized ten years ago in 1996. The basic questions were the distribution of homogenisation methods and the overall use of homogenous (homogenised) time series in climate change studies in that time. Homogenisation was not widely accepted, and the generally recommended methods had very simple and poor mathematical basis.

The general view has been changed since then. Homogenisation became a basic element of the data quality procedure, although many of the recommendations of the First Seminar are not fulfilled even today. The information on the applied homogenisation method is not always published along with the time series, however we have many good examples already.

At the same time, research community requires more and more from the experts on homogenisation. Nowadays, one of the largest, still not fully solved problems is the homogenisation of daily time series. Many, very important indices are calculated from the daily data, and those indices are needed for climate change detection, changes of extreme values, etc.

A new COST Action proposal has the basic task to compare, evaluate and develop homogenisation methods. COST (European Cooperation in the field of Scientific and Technical Research) is open for the appropriate institutions, but supports only the member states. We hope, that this new Action will give a push to the dissemination and development of homogenisation methods not only in Europe, but worldwide.

This seminar is supported by WMO and OMSZ, and we hope that the series of Homogenisation Seminars can co-operate with different other initiative for development of data quality with special regards to homogenisation.

Sándor Szalai

CONTENTS

AN OVERVIEW ON THE MAIN METHODOLOGICAL QUESTIONS OF HOMOGENIZATION

Tamás Szentimrey

Hungarian Meteorological Service
szentimrey.t@met.hu

1. INTRODUCTION

Today the theme of homogenization can be divided into two subgroups, such as monthly and daily data series homogenization. These subjects are in strong connection of course, for example monthly results can be used for the homogenization of daily data. However while monthly series homogenization is relatively well elaborated, the problem of daily data is still in early stage. Owing to these respects we will consider first the main methodological questions of monthly series homogenization, namely the relative and absolute test methods. In connection with the relative methods, the following topics will be detailed: mathematical methodology for comparison of series, break point (changepoint) and outlier detection, correction of series, missing data complementing as well as the possibilities of verification for both methods and results. Concerning the homogenization of daily data series we will discuss the possibility to use the detected monthly inhomogeneities for daily data, furthermore the special importance of quality control and missing data completion. The key issue of exact mathematical methods is also emphasized, as there is no royal road to anything.
The following methods will be referred as examples: SNHT (Standard Normal Homogeneity Test, Alexandersson, 1986), Caussinus-Mestre's method (Mestre, 1999, 2004) and MASH (Multiple Analysis of Series for Homogenization, Szentimrey, 1999, 2004).

2. HOMOGENIZATION OF MONTHLY DATA

2.1 Absolute and relative methods

In case of absolute methods we have only one candidate series without any other reference series. The additive model for the candidate monthly series is,
$$X(t) = \mu(t) + E + IH(t) + \varepsilon(t) \qquad \left(t = 1,2,\ldots,n\right) ,$$
where $\mu(t)$ is the unknown climate change signal, E is the spatial expected value, $IH(t)$ is the inhomogeneity signal and $\varepsilon(t)$ is a normal white noise series. The main problem of the application of absolute methods is that the separation between the climate change signal and the inhomogeneity is essentially impossible.

2.1.1. The additive model of relative methods

Relative methods can be applied if there are more station monthly series given, which can be compared mutually. In case of relative methods the additive model for more monthly series belongs to the same month in a small climate region is as follows,
$$X_j(t) = \mu(t) + E_j + IH_j(t) + \varepsilon_j(t) \qquad \left(j = 1,2,\ldots,N;\ t = 1,2,\ldots,n\right) , \qquad (1)$$

where $\mu(t)$ is the common and unknown climate change signal, E_j are the spatial expected values, $IH_j(t)$ are the inhomogeneity signals and $\varepsilon(t)$ are normal white noise series. As concerns the type of $\mu(t)$ there is no assumption about the shape of this signal. The type of inhomogeneity $IH(t)$ is in general a 'step-like function' with unknown break points T and shifts $IH(T)-IH(T+1)\neq 0$, and $IH(n)=0$ is assumed in general. The expected values

$$\mathrm{E}(X_j(t))=\mu(t)+E_j+IH_j(t) \qquad (j=1,2,\ldots,N;\ t=1,2,\ldots,n)\ ,$$

are covered with the normal white noise series,

$$\boldsymbol{\varepsilon}(t)=[\varepsilon_1(t),\ldots\ldots,\varepsilon_N(t)]^{\mathrm{T}}\in N(\mathbf{0},\mathbf{C}) \qquad (t=1,\ldots,n)\ ,$$

where the vector variables $\boldsymbol{\varepsilon}(t)$ $(t=1,\ldots,n)$ are totally independent in time, and matrix $\mathbf{C}$ is the spatial covariance matrix between the stations. This station covariance matrix $\mathbf{C}$ may have a key role in methodology of comparison of series.

The aim of the homogenization procedure is to detect the inhomogeneities and to correct the series. During the procedure the series can be compared mutually and the role of series – either the candidate or the reference ones – is changing in the course of procedure. The reference series are not assumed to be homogeneous in the correct examinations! The significance and the power of the procedures can be defined according to the probabilities of the type of errors. Type one error means the detection of false or superfluous inhomogeneity while type two error means neglecting some real inhomogeneity.

2.2 Methodology for comparison of series

The problem of comparison of series is related to the following questions: reference series creation, difference series constitution, multiple comparisons of series etc. This topic is very important for detection as well as for correction, because the efficient comparison of series can increase both the significance and the power. The development of efficient comparison methods can be based on the examination of the spatial covariance structure of data series.

As we emphasized earlier all the examined series $X_j(t)$ $(j=1,\ldots,N)$ are taken as candidate and reference series alike, besides the reference series are not assumed to be homogeneous at the correct examinations!

The main problem arises from the fact that the shape of climate change signal is unknown. Therefore so-called difference series are examined in order to filter out the climate change signal $\mu(t)$. The simple difference series between pairs are $Z(t)=X_j(t)-X_i(t)$. However the difference series constitution can be formulated in more general way as well. Assuming that $X_j(t)$ is the candidate series and the other ones are the reference series, the difference series belongs to the candidate series can be constituted as,

$$Z_j(t)=X_j(t)-\sum_{i\neq j}\lambda_{ji}X_i(t)=IH_j(t)-\sum_{i\neq j}\lambda_{ji}IH_i(t)+\varepsilon_{Z_j}(t) \qquad (2)$$

with condition of $\sum_{i\neq j}\lambda_{ji}=1$ for the weighting factors. As a result of the last condition, the unknown climate change signal $\mu(t)$ has been filtered out. Consequently the inhomogeneities can be detected by the examination of the difference series defined according to formula (1). The interpolation series $\sum_{i\neq j}\lambda_{ji}X_i(t)$ can be taken as created reference series for candidate series $X_j(t)$.

In addition if we want to increase the signal to noise ratio in order to increase the power of detection then we have to minimize the variance of noise term $\varepsilon_{Z_j}(t)$.

The covariance matrix **C** uniquely determines the optimum weighting factors that minimize the variance, and the optimal difference series created in this manner can be applied efficiently for the detection and correction procedures (MASH, Szentimrey, 1999). We mention that in case of using the generalized-least-squares estimation for the unknown climate change signal $\mu(t)$, also the optimal difference series is obtained with minimal variance. We have to examine more difference series in order to separate the appropriate detected inhomogeneities for the candidate series. More difference series created without common reference series and with minimal variances can be defined as optimal difference series system (MASH).

2.3 Methodology for break point (changepoint) detection

One of the basic tasks of the homogenization is the examination of (more) difference series in order to detect the break points and to separate them for the candidate series.
Let $Z(t)$ be a difference series according to the formula (2), that is

$$Z(t) = IH_Z(t) + \varepsilon_Z(t) \qquad (t = 1,..,n), \tag{3}$$

where $IH_Z(t)$ is a mixed inhomogeneity of difference series $Z(t)$ with K break points $T_1 < T_2 < ..... < T_K$. In general the number K and the position of the multiple break points $T_1 < T_2 < ..... < T_K$ are unknown, furthermore the noise variables $\varepsilon_Z(t) \in N(E_Z, \sigma_Z^2)$ $(t = 1,...,n)$ are totally independent in time. The basic types of the detection procedures are the stepwise and the multiple break points detection. Let us have the following notation of the estimates: $\hat{K}$; $\hat{T}_1 < \hat{T}_2 < ..... < \hat{T}_{\hat{K}}$

2.3.1 Stepwise break points detection procedures

The algorithm of the stepwise decision procedure to detect the break points is as follows.
Step 1: the 'most probable' break point $\hat{T}_1^{(1)}$.
Step 2: the 'most probable' second break point $\hat{T}_2^{(2)}$, assuming that $\hat{T}_1^{(1)}$ is a real break point. …………………………………..
Step $\hat{K}$: the 'most probable' $\hat{K}^{th}$ break point $\hat{T}_{\hat{K}}^{(\hat{K})}$, assuming that $\hat{T}_1^{(1)},..,\hat{T}_{\hat{K}-1}^{(\hat{K}-1)}$ are real break points.
The number $\hat{K}$ is the estimation for the number of break points, which is determined also in the course of procedure.
As regards the concept of 'most probable', it depends on the aim, there is no absolute objective function. In general the maximum likelihood estimation is applied.
The method SNHT (Standard Normal Homogeneity Test, Alexandersson, 1986) is an example for the application of this stepwise principle for break points detection. However the multiple break points detection procedures, when the break points are estimated jointly instead of step by step, are more exact and elegant than the stepwise ones in mathematical respect.

2.3.2 Multiple break points detection procedures

For joint estimation of the break points there are different possibilities, principles, which are classical ways in mathematical statistics.

2.3.2.1 Detection based on Bayesian Approach

The methods based on Bayesian model selection are the penalized likelihood methods. These methods are different in the penalty terms or criterions e.g. Akaike criterion, Schwarz criterion, Caussinus-Lyazrhi criterion.
The Caussinus-Mestre's procedure (Mestre, 1999, 2004) based on the Caussinus-Lyazrhi criterion is an example for the penalized likelihood methods.

2.3.2.2 Detection based on Test of Hypothesis

Another possibility is to use hypothesis test methods for the detection of break points. At the MASH method (Szentimrey, 1999) a hypothesis test procedure has been developed, as we want to avoid the type one error that is the damage of data series. The essence of this multiple break points detection procedure based on test of hypothesis on a given significance level is as follows.
If the detected break points of $Z(t)$ are $\hat{K};\ \hat{T}_1 < \hat{T}_2 < ..... < \hat{T}_{\hat{K}}$,
then on the given significance level p (e.g.: p=0.1):

i, $Z(t)$ is not homogeneous above the intervals $(\hat{T}_{k\text{-}1}, \hat{T}_{k+1}]$ because,

$$\mathrm{P}\left(\ \exists(\hat{T}_{k\text{-}1}, \hat{T}_{k+1}]\ \text{above that}: Z(t)\ \text{homogeneous}\ \right) = p$$

Consequently the detected break points $\hat{T}_k$ are not superfluous.

This means there is no serious type one error.

ii, $Z(t)$ can be accepted to be homogeneous above the intervals $(\hat{T}_{k\text{-}1}, \hat{T}_k]$.

This means there is no serious type two error.

Remark
Confidence intervals are also given for the break points beside the point estimates at the method MASH (Szentimrey, 1999).

2.3.3 Outlier detection (QC)

In case of monthly series homogenization the outlier detection is the quality control (QC) procedure for the data. Furthermore the outlier detection can be considered as a special part of break points detection, because an outlier is equivalent with two special break points. The point T_{out} is an outlier point if and only if

$$IH(T_{out}) - IH(T_{out} - 1) = IH(T_{out}) - IH(T_{out} + 1) \neq 0 .$$

Consequently $T_{out} - 1$ and T_{out} are break points, where their shifts are the same in absolute value but with opposite sign.

2.4 Methodology for correction of series

Beside the detection, another basic task of the homogenization is the correction of series. Calculation of correction factors can be based on the examination of difference series for estimation of shifts $IH(\hat{T}_k) - IH(\hat{T}_k + 1)$ $\left(k = 1,...,\hat{K}\right)$ at the detected break points.

2.4.1 Correction methods

Almost all the methods use point estimation for the correction factors at the detected break points. For example the Caussinus and Mestre's method (Mestre, 2004) uses the standard least squares technique to estimate the correction factors.

The MASH procedure (Szentimrey, 1999) is an exception because the correction factors are estimated on the basis of confidence intervals.

2.4.2 Missing data complementing

In fact the missing data completion or filling the gaps is an interpolation problem. At the MASH method (Szentimrey, 1999) the applied spatial interpolation formula is in accordance with the series comparison formula (2), i.e.

$$\hat{X}_j(t) = \lambda_0 + \sum_{i \neq j} \lambda_{ji} X_i(t) \tag{4}$$

where $\hat{X}_j(t)$ is the interpolated candidate series and the series $X_i(t)\,(i \neq j)$ are the reference ones, with condition of $\sum_{i \neq j} \lambda_{ji} = 1$ for the weighting factors. The optimum interpolation parameters that minimize $RMSE$ error are uniquely determined by the covariance matrix $\mathbf{C}$.

2.5 The philosophy of MASH

Careful break points detection and correction iteration procedures in order to decrease the probability of type one error. At the same time using optimal series comparison for decreasing the probability of type two error i.e to increase the power.
The break points detection is based on hypothesis testing, point estimation and confidence intervals.
The correction is also based on point estimation and confidence intervals.
Series comparison uses optimal difference series constitution with optimal weighting factors.
Missing values are completed by spatial interpolation with optimal weighting factors.
In addition the software MASH is an iteration procedure (Szentimrey, 2006)!

2.6 Possibilities of Verification

The confidence in the homogenized series may be increased by the examination of both the methods and the results.
Possibilities for the examination of methods:

- ⇒ Theoretical overview and evaluation of the homogenization methods.
- ⇒ Testing the methods on the basis of artificial, generated series.

Possibilities for the examination of homogenization results:

- ⇒ Comparing the detected inhomogeneities with the Meta Data.
- ⇒ Mathematical verification procedures to evaluate the results.

BASIC CONCEPTION OF VERIFICATION PROCEDURE BUILT IN MASH (SZENTIMREY, 2004):

The quality of the homogenized series can be evaluated by the joint comparative mathematical examination of the original and the homogenized series systems.
Also Meta Data information can be used and tested during the procedure.
Finally, as regards testing of methods we emphasize that,
Homogenization $\neq$ Break Points Detection!!!
Homogenization is a much more complex problem:
Homogenization $=$Comparison+ Detection+Correction+etc.
If we want to test the methods, we must fully aware of the complexity!

3. HOMOGENIZATION OF DAILY DATA

The main question is the relation of daily and monthly homogenization.
The alternative possibilities are as follows:
– To use the detected monthly inhomogeneities directly for daily data homogenization.
– Direct methods for daily data homogenization.
The problems connected with the possibilities:
– The direct use of the detected monthly inhomogeneities is probably not sufficient.
– Direct methods for daily data homogenization is probably not enough efficient thinking of the larger variability (less signal to noise ratio).
So we have the following question:
How can we use the valuable information of detected monthly inhomogeneities for daily data homogenization?

REFERENCES

Alexandersson, H., 1986: „A homogeneity test applied to precipitation data”, Int. J. Climatology 6: 661-675.

Mestre, O., 1999: „Step-by-step procedures for choosing a model with change-points”, Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 15-25.

Mestre, O., 2004: „Detecting multiple change-points in a gaussian sample using dynamic programming”, Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary; WMO, WCDMP-No. 56, pp. 89-92.

Mestre, O., 2004: „Correction climate series using ANOVA technique”, Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary; WMO, WCDMP-No. 56, pp. 93-96.

Szentimrey, T., 1999: „Multiple Analysis of Series for Homogenization (MASH)”, Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2004: „Multiple Analysis of Series for Homogenization (MASH); Verification procedure for homogenized time series”, Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary; WMO, WCDMP-No. 56, pp. 193-201.

Szentimrey, T., 2006: „Manual of homogenization software MASHv3.01”, Hungarian Meteorological Service, p. 61.

Changepoint Detection in Periodic and Autocorrelated Time Series

Robert Lund[1], Xiaolan L. Wang[2], Jaxk Reeves[3]
Yang Feng[2], QiQi Lu[4], Colin Gallagher[1]

1. Department of Mathematical Sciences, Clemson University, Clemson, SC;
   Email: lund@ces.clemson.edu
2. Climate Research Division, ASTD, STB, Environment Canada, Toronto, ON
3. Department of Statistics, The University of Georgia, Athens, GA
4. Department of Mathematics and Statistics, Mississippi State University, MS

(*Journal of Climate*, accepted subject to revision, June 2006)

ABSTRACT

Undocumented changepoints (inhomogeneities) are ubiquitous features of climatic time series. Level shifts in time series caused by changepoints confound many inference problems and are very important data features. Tests for undocumented changepoints from models that have independent and identically distributed errors are by now well understood. However, most climate series exhibit serial autocorrelation. Series recorded at monthly, daily, or hourly frequencies may also have periodic structures. This article develops a formal statistical test for undocumented changepoints for periodic and autocorrelated time series. Classical changepoint tests based on sum of squared errors are modified to take into account series autocorrelations and periodicities. The methods are applied in the analyses of a monthly pressure and a monthly temperature series.

1. INTRODUCTION

A changepoint in a time series is a time at which the structural pattern of the series changes. This shift is typically measured in terms of mean or average levels, but changepoints in variability, or more generally, in the marginal distributions of the series, could be studied. The series under study may contain measurements scaled by a reference series or the raw observations themselves. While such distinctions are not crucial for the moment, the reader is referred to Alexandersson (1986) for a discussion on comparisons made by forming differences and ratios of target and reference series.

Instrumentation/observer changes, station location changes, and changes in observation practices are frequent non-climatic (artificial) culprits behind changepoints. In many cases, the changepoint time and cause are documented and it is reasonably straightforward to statistically adjust (homogenize) the series for the effects of the changepoint. Unfortunately, many changepoints are undocumented.

While undocumented changepoints are sometimes evident in a plot of the series, debatable cases also abound. Visual detection of a changepoint in a series with a prominent seasonal mean can be difficult (Section 5b gives an example). Moreover, the statistical methods used to identify undocumented changepoints are known to be very important (see Lund and Reeves 2002, Reeves et al.~2006, and Wang 2006). Undocumented changepoint detection methods can greatly reduce the workload of metadata investigation by identifying times around which the investigation should focus. Hence, the development of statistically sound tests for undocumented changepoints is desirable. Undocumented changepoint detection in climate settings has been previously explored by Potter (1981), Thompson (1984), Alexandersson (1986), Solow (1987), Karl and Williams (1987), Gullet et al.~(1991), Rhoades and Salinger (1993), Easterling and Peterson (1995), Alexandersson and Moberg (1997), Vincent (1998), Lund and Reeves2002), Ducre-Robitaille et al.~(2003), Wang (2003), Wang and Feng (2004), Hanesiak and Wang (2005), and Wang (2006). The statistical side of the subject is also vast, with Page (1955), Kander and Zacks (1966), Hinkley (1969 and 1971), Brown et al. (1975), Hawkins (1977), Chen and Gupta (2000), and Caussinus and Mestre (2004) being a prominent sample. Neither of these lists is complete. Reeves et al.~(2006) review undocumented changepoint detection methods in climate settings for models with independent and identically distributed (IID) Gaussian errors.

In this paper, we develop a method for undocumented changepoint detection for series with autocorrelated and periodic features. The periodic and autocorrelation aspects are modeled in tandem rather than separately. The results enable one to test for undocumented changepoints in a variety of realistic climate settings. The methods work with or without a reference series and are specifically designed to handle correlated series. They greatly alleviate the heavy dependence of most existing methods (which assume IID Gaussian errors) on the availability and use of good homogeneous reference series to diminish the effects of periodicities and autocorrelations (and trends for models without a trend term, e.g., Alexandersson 1986). This paper is perhaps the first detailed investigation of changepoint detection in climate settings involving autocorrelation; periodicities and changepoint detection were previously considered in Gullett et al. (1991).

The rest of this paper proceeds as follows. Section 2 introduces a time series regression model with autocorrelated and periodic features. A test statistic weighing a null hypothesis of overall series homogeneity against the alternative of an undocumented changepoint is developed in Section 3. Section 4 shows why autocorrelation and periodicities are important in changepoint detection problems. Applications of the methods to monthly pressures and monthly temperatures are given in Section 5.

2. THE MODEL

In the time homogeneous (non-periodic) setting, a simple but useful model allowing for one changepoint in time series X_t is the following regression:

$$X_t = m + bt + \Delta 1_{[t>c]} + e_t,\ 1 \le t \le N, \qquad (1)$$

where c is the unknown time of change (the exposition focuses on detection of one changepoint but others could exist in practice), the magnitude of the changepoint effect (step-size) is Δ, and e_t is a zero mean random variable that may be autocorrelated (a time series). The factor b allows for a simple linear trend in series values. Following Wang (2003), the trend (slope) b is constrained as equal before and after the changepoint at time c. The simple linear structure in (1) may require modification in some settings; for example, Lund and Reeves (2002) analyze a carbon dioxide series where a quadratic trend is apparent, while the existence of a good reference series may render the inclusion of a trend component unnecessary.

Equation (1) is a simple linear regression model with two phases; such models and their variants have been studied in Hinkley (1969), Solow (1987), Easterling and Peterson (1995), Vincent (1998), Lund and Reeves (2002), and Wang (2003). A periodic variant of (1) merely allows the location parameter m to depend on time and vary periodically with period T, i.e.,

$$X_{nT+v} = m_v + b(nT+v) + \Delta 1_{[nT+v>c]} + e_{nT+v},\ 1 \le nT+v \le N. \qquad (2)$$

In (2), X_{nT+v} refers to the series during the v season (or month or day…) of cycle n. The seasonal index v satisfies $1 \le v \le T$ and the period T is assumed known. Our bookkeeping assumes d complete cycles of data and labels these cycles as 0, …, d-1, respectively; this makes X_1 the observation for season 1 of cycle 0. The total number of observations is $N = dT$. Equation (2) assumes a time homogeneous (non-periodic) linear trend and time homogeneous mean-shift; this is emphasized notationally by the fact that b and Δ do not carry subscripts of v. Changepoints inducing different effects on different seasons could be modeled by allowing Δ to depend on v, but we will not pursue such generality here.

The mean series response at time $(nT+v)$ in (2) is

$$E(X_{nT+v}) = m_v + b(nT+v) + \Delta 1_{[nT+v>c]};$$

hence, seasonality in the first moment is described by. In addition to seasonality in mean, many climatic series also display seasonality in variance and autocorrelations. For examples, non-tropical temperature series show larger variabilities (lag zero autocovariances) during winter seasons and many Western United States precipitation series have minimal variability during late summer and early fall seasons. Changepoint times are slightly more difficult to detect at times of peak series variability (see Section 4). To allow for autocorrelation and periodicities, the regression errors e_t are modeled as a periodically stationary time series (periodic series). A general overview of periodic series and their applications in climate modeling is presented in Lund et al. (1995).

For simplicity of computation, presentation, and flexibility, we will work with perhaps the simplest periodic time series model for e_t: a first order periodic autoregression, PAR(1). Such an e_t is governed by the periodic difference equation:

$$e_{nT+v} = f_v e_{nT+v-1} + Z_{nT+v} \qquad (3)$$

where $Z_t = Z_{nT+v}$ is mean zero periodic white noise; that is, Z_t and Z_s are uncorrelated when $t \neq s$, Z_t has zero mean for every t, and the variance of Z_t is periodic in that $Var(Z_{nT+v}) = s_v^2$. The model in (3) has $2T$ parameters --- this total may be large if the series is observed frequently. For example, a daily PAR(1) ($T = 365$) has 365 autoregressive parameters and \$365\$ white noise variance parameters. Parsimony issues for periodic series are discussed in Lund et al. (2006).

3. THE TEST STATISTIC

An undocumented changepoint test statistic weighs the null hypothesis that $\Delta = 0$ (termed a null or H_0 model) against the alternative that $\Delta \neq 0$ (termed a full or H_A model). The changepoint time c is an unknown parameter of the full model. The general form of the test statistic coincides with that in Lund and Reeves (2002) and Wang (2003):

$$F_{\max} = \max_{1 \le c < N} F_c \qquad (4)$$

where F_c defined by

$$F_c = \frac{SSE_0 - SSE_A(c)}{SSE_A(c)/(N-p)} \qquad (5)$$

is a regression *F*-type statistic measuring closeness of the null model and a full model with a single changepoint at time *c*. In (5), SSE_0 and $SSE_A(c)$ are null model sum of squared errors and full model sum of squared errors when a changepoint exists at time *c*, respectively. Note that SSE_0 does not depend on the value of *c* but that

$SSE_A(c)$ does. Here, p is the number of free parameters involved in the full model with a changepoint at time c; later, we present applications where $p = 8$.

In classical regression settings with IID Gaussian e_t, F_c has an F distribution (exactly) with 1 numerator degree of freedom and $(N - p)$ denominator degrees of freedom. The larger F_c is, the more evidence there is that an undocumented changepoint exists at time
c. Intuitively, the $F_{\max}$ statistic selects the time of largest discrepancy in the two phases of the model, as measured by regression F_c statistics, as the estimator of c. The null hypothesis H_0 is accepted when $F_{\max}$ is small enough to be explained by chance variation and rejected when $F_{\max}$ is excessively large. The null hypothesis percentiles of the $F_{\max}$ distribution, assuming IID Gaussian errors (specifically time-homogeneity and independence) and the regression form in (1) are tabulated in Wang (2003). The reader is cautioned about historical mistakes in quantifying this distribution (see Lund and Reeves 2002 for discussion); the percentiles in Wang (2003) and Lund and Reeves (2002) are accurate.

For IID e_t, Alexandersson (1986) and Lund and Reeves (2002) connect the $F_{\max}$ statistic to Gaussian likelihood ratios and maxima of correlated t and F random variates. Since the F_c's are correlated in c, $F_{\max}$ does not behave statistically as the maximum of independent F-statistics, with each F-statistic having 1 numerator and $(N - p)$ denominator degrees of freedom.

The key methodological innovation put forth here involves modifying sums of squares in autocorrelated (and periodic) settings. Here, we use weighted squared prediction errors:

$$SSE_0 = \sum_{n=0}^{n} \sum_{v=1}^{T} \frac{(X_{nT+v} - \hat{X}^0_{nT+v})^2}{s_v^2}, \qquad (6)$$

and

$$SSE_A(c) = \sum_{n=0}^{n} \sum_{v=1}^{T} \frac{[X_{nT+v} - \hat{X}^A_{nT+v}(c)]^2}{s_v^2}; \qquad (7)$$

moreover, the form of the predictions $\hat{X}^0_{nT+v}$ and $\hat{X}^A_{nT+v}(c)$ now become best one-step-ahead linear predictions:

$$\hat{X}^0_{nT+v} = P_0[X_{nT+v} \mid X_1, ..., X_{nT+v-1}, 1], \ \hat{X}^A_{nT+v} = P_A[X_{nT+v} \mid X_1, ..., X_{nT+v-1}, 1], \quad (8)$$

where $P[X \mid Y, 1]$ denotes the best (minimum mean square error) linear prediction of X from linear combinations of Y and a constant. The subscript under P (or the superscript on $\hat{X}_t$ indicates the model (null or full) under which the linear prediction is to be computed. Brockwell and Davis (1991, Chapter 8) provide the theory for sum of squared errors in time series settings. We comment that the PAR(1) structure gives $Var[X_{nT+v} - \hat{X}_{nT+v}] = s_v^2$.

The computation of $F_{\max}$ requires SSE_0 and $SSE_A(c)$ for each c. We first tackle SSE_0. For the null model, the PAR(1) structure gives

$$\hat{X}^0_{nT+v} = m_v + b(nT+v) + f_v[X_{nT+v-1} - m_{v-1} - b(nT+v-1)] \qquad (9)$$

for $2 \le nT+v \le N$, where the startup convention $\hat{X}^0_1 = m_1 + b$ is made. As the parameters b, m_v and f_v $(1 \le v \le T)$ are unknown, we estimate them by numerically minimizing the sum of squares over all feasible values. Since it is statistically wasteful to expend *T* parameters each in modeling m_v and f_v, we impose the first-order Fourier parsimony constraints

$$m_v = A_0 + A_1 \cos\left(\frac{2p(v-t)}{T}\right), \quad f_v = B_0 + B_1 \cos\left(\frac{2p(v-h)}{T}\right), \qquad (10)$$

upon model parameters during minimization. In these formulations, A_0 and B_0 are the mean periodic value of the parameter being modeled, with A_1 and B_1 representing the maximum amplitude above or below the mean which could be achieved. The t and h are the times (phases) in the cycle at which the parameter being modeled achieves its maximum. These phase parameters are unique only modulo *T*, but it is conventional to utilize values in the range [0, *T*]. Higher order Fourier series and/or wavelet based expansions could be considered if needed. Lund et al. (2006) discuss periodic parsimonious time series modeling in general. Notice that the null model expends six free parameters in modeling m_v and f_v. Adding the trend parameter b brings the free parameter count in the expression for $\hat{X}^0_{nT+v}$ to seven.

The parameters s_v^2 are viewed here as nuisance parameters. These parameters are least squares weights and become increasingly important to know accurately when the variance of the time series has a large seasonal cycle. In practice, one needs only a rough idea of their values --- and this is easily accomplished by several methods, one of which is presented in the applications in the next section.

To compute $SSE_A(c)$ for a fixed *c*, we proceed as with the null model except that (9) is modified to account for the changepoint at time *c*:

$$\hat{X}^{0}_{nT+v} = m_v + b(nT+v) + \Delta 1_{[nT+v>c]}$$
$$+ f_v[X_{nT+v-1} - m_{v-1} - b(nT+v-1) - \Delta 1_{[nT+v-1>c]}] \quad (11)$$

for $2 \le nT + v \le N$, where the startup convention $\hat{X}^{A}_{1} = m_1 + b$ is made. The full model involves eight free parameters (add one for Δ) in describing $\hat{X}^{A}_{nT+v}$ for each fixed *c*, one more than the null model's total.

The $F_{\max}$ statistic in (4) along with SSE_0 and $SSE_A(c)$ defined in (6) and (7) has, approximately, the percentiles reported in Wang (2003) with $(N-p)$ denominator degrees of freedom. These percentiles are not exact as some of the parameters in the fitted time series model must be estimated; however, as the sample size *N* increases, the percentiles become increasingly accurate. Overall, perturbations in Wang's (2003) percentiles induced by parameter estimation are relatively minor when compared to those reported in the next section when autocorrelation features of the series are ignored.

4. THE IMPORTANCE OF AUTOCORRELATION AND PERIODICITIES

This section shows how autocorrelations and periodicities influence changepoint detection procedures. We will study how the 95th percentile of the $F_{\max}$ distribution in (4) changes under correlation and periodic time series features. Our intent here is to mimic what happens when one ignores the autocorrelations and/or periodicities in the series.

We first investigate the effects of autocorrelation. Table 1 displays the sample 95th percentile of the $F_{\max}$ distribution under various levels of serial autocorrelation as governed by a first-order autoregression, AR(1). To isolate the effects of autocorrelation only, we examine the time-homogeneous model in (1) with AR(1) e_t satisfying

$$e_t = f e_{t-1} + Z_t, \quad (12)$$

where $\{Z_t\}$ is IID Gaussian noise with variance s^2. Equation (12) is merely a time homogeneous version of (3). We consider series of length *N* = 100. As the autoregressive coefficient f increases, the degree of serial autocorrelation in the model increases. The white noise variance s^2 is selected to make the variance of the error series e_t unity in all cases; this allows for meaningful comparisons across table entries. The estimated 95th percentiles were aggregated from one hundred thousand independent simulations each; hence, they are reasonably accurate.

Table 1. The 95th percentiles of the $F_{\max}$ statistic under autocorrelation f.

f	-0.95	-0.75	-0.50	-0.25	0.00	0.15	0.25	0.50	0.75	0.95
$F_{\max,0.95}$	4.531	4.211	5.323	7.460	11.054	14.350	17.250	29.547	64.185	176.753

The results show that the 95th $F_{\max}$ percentile generally increases with increasing f, dramatically so for values of f slightly less than unity. This agrees with the findings of Percival and Rothrock (2005) and is not surprising: as $f > 0$ becomes larger, the series makes longer sojourns above and below its mean response values, which effectively imitates the effects of a mean-shift due to a changepoint. Hence, one should be cautious when declaring changepoints in positively correlated series. When $f < 0$ (which is not usually realistic in climate modeling), consecutive observations tend to split the mean response level (one above and one below) and make changepoints easier to detect. The case where $f = 0$ corresponds to the case of independent errors and is considered in detail by Wang (2003).

Next, we consider how periodicities in the white noise variances influence changepoint detection. Here, our model is (2) and $\{e_t\}$ is periodic Gaussian white noise. To isolate on the effects of variance-periodicities only, we take $m_v \equiv 0$, $f_v \equiv 0$, and $b = 0$. The white noise variance parameters were assumed sinusoidal:

$$s_v^2 = C_0 + C_1 \cos\left(\frac{2p(v-x)}{T}\right). \qquad (13)$$

Table 2 reports estimates of the 95th $F_{\max}$ percentile for the regression model in Wang (2003), again computed for one hundred thousand simulations for each entry, for various values of C_0 and C_1 (and $x = 0$). If $C_1 = 0$, the error variances are nonseasonal and the setting reduces to that studied in Wang (2003). In fact, when $C_1 = 0$, the $F_{\max}$ percentiles do not depend on the value of C_0. The larger C_1 is relative to C_0, the more seasonality there is in white noise variances (as measured across varying seasons *v*). Of course, we take $C_1 < C_0$ or s_v^2 could be negative. This table employs *N* = 120 (*d* = 10, *T* = 12), which corresponds to a decade of monthly data.

The Table 2 results show that the null hypothesis $F_{\max}$ percentiles increase slightly with increasing seasonal variability. This is as expected: when s_v^2 varies greatly with the season *v*, there is a larger chance for an outlying e_t to pull the least squares regression fit away from its true zero mean baseline, hence mimicking a changepoint mean-shift. Note, however, that the effects of variance-seasonality are nowhere near as drastic as those of autocorrelation. Observe that the percentiles for the first three entries are approximately those for the second three entries. Those for the case $C_1 = 0$ indeed coincide with those reported in Wang (2003) (up to simulation error).

Table2. The 95th percentiles of the $F_{\max}$ statistic under variance-seasonality.

C_0	C_1	$F_{\max,0.95}$
1.00	0.00	11.107
1.00	0.50	11.926
1.00	0.95	13.253
10.0	0.00	11.093
10.0	5.00	11.926
10.0	9.50	13.254

Finally, to obtain some feel for practical cases involving both periodic and autocorrelated features, we have simulated a case where the $\{e_t\}$ follow (3) and where the f_v and s_v^2 parameters are set to the values that are fitted in the temperatures series analyzed in Section 5b below. Here, the series length is N = 600. A simulation of one hundred thousand runs gives an estimated 95th percentile of about 16.16, about 40% larger than that for IID errors (which is 11.55, as tabulated in Wang 2003). Hence, even in practical cases, there is plenty of room to commit mistakes by ignoring autocorrelation and periodicities.

5. EXAMPLES

a. A monthly mean atmospheric pressure series

The above methods were applied to a series of monthly mean atmospheric pressures recorded at Stephenville Airport (Newfoundland, Canada) for the period 1953-2002 (50 years; N = 600), shown in Figure 1. The mean pressure is lower in winter (December-March) and peaks in summer; winter pressure variabilities are also higher than those in summer, a property also shared by temperatures.

To estimate s_v^2, a null model and full models for each admissible changepoint time c were fitted to the data by numerically minimizing the sum of squared errors in (6) and (7), *without the weights* s_v^2 *in the denominator of these equations.* The numerical minimization was a fairly stable endeavor and was accomplished with a gradient step and search algorithm. The fits suggest that c = 286 is the most likely time of a changepoint. The parameters s_v^2 are now simply estimated as the sample variance of the season v residuals $(X_{nT+v} - \hat{X}_{nT+v}^A)$ computed from a full model fit with c = 286:

$$\hat{s}_v^2 = \frac{1}{d}\sum_{n=0}^{d-1}(X_{nT+v} - \hat{X}_{nT+v}^A)^2 .$$

Figure 2 plots these estimates and confirms that winter months are the most variable. We may now compute F_c and $F_{\max}$, accounting for the effects of autocorrelations and variance-periodicities.

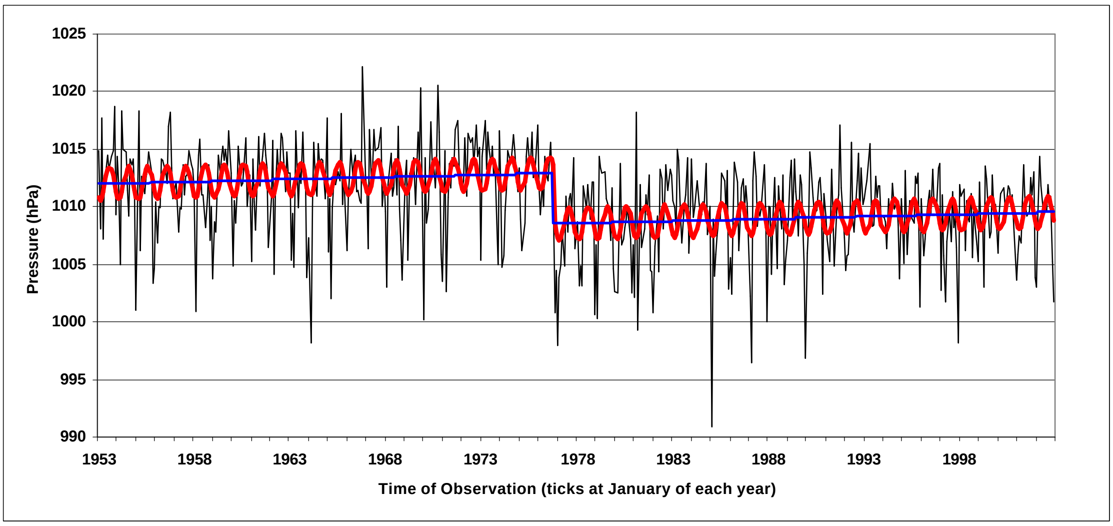


Figure 1. Monthly mean atmospheric pressures recorded at Stephenville Airport (Newfoundland, Canada). The red curve is the fitted regression response, the blue line, the trend.

Figure 3 plots values of F_c against a 95% confidence threshold constructed assuming a null hypothesis of no changepoint. The largest F_c is $F_{\max} = 71.075$, which occurred at c = 286 (October 1976) and greatly exceeds the 95% threshold. Hence, evidence suggests an extremely significant changepoint around October of 1976. The mean-shift at the changepoint is estimated as $\hat{\Delta} = -4.4$ hPa by the full model and the trend estimate is $\hat{b} = 0.00324$. The other parameters in the full model fit for c = 286 are $\hat{A}_0 = 1011.968$, $\hat{A}_1 = 1.404$, $\hat{t} = 8.005$, $\hat{B}_0 = 0.0637$, $\hat{B}_1 = 0.0422$, and $\hat{h} = 8.926$.

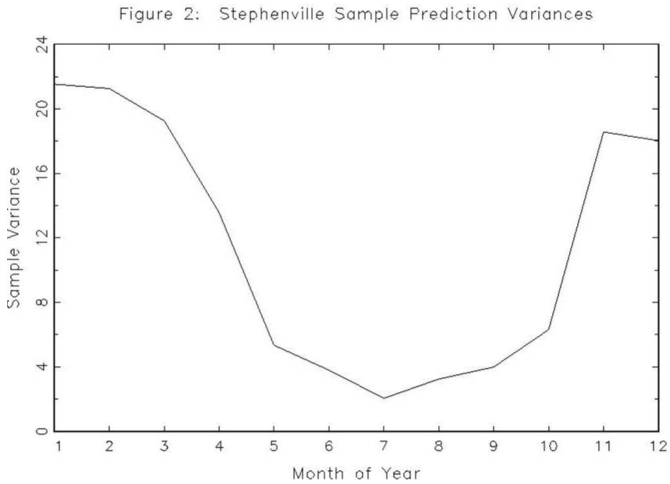


Figure 2. Sample prediction variances of the Stephenville pressure series.

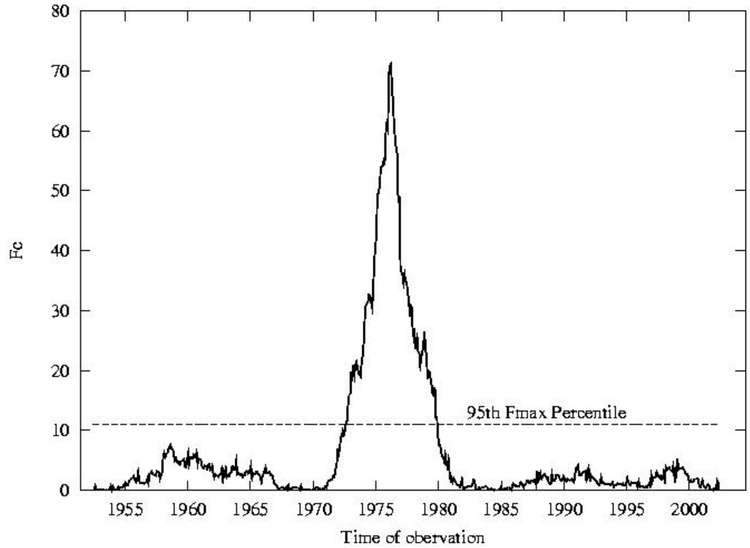


Figure 3. The F_c statistics of the Stephenville pressure series.

Apparently, the changepoint was caused by neglecting the 25.6 m station elevation in the calculation of station pressures from barometer readings prior to 1977 (i.e., an elevation of 0 m was used instead of 25.6 m). According to a physically based estimate using a hydrostatic model and hourly pressure and temperature data (see Wan et al. 2006), neglecting such an elevation causes a bias of 3.2 hPa on pressure values. The estimated changepoint time is very close to its true value. Additional changes that happened between December 1976 and January 1977, such as the use of computer-produced pressure reduction tables and the addition of a plateau correction, may have also contributed to the magnitude of the mean-shift. This is likely why the magnitude of $\hat{\Delta}$ exceeds 3.2 hPa.

This was not a hard changepoint to identify. In fact, the two-phase regression approach of Wang (2003) for time-homogeneous data also identifies a drop of 4.2 hPa between September and October 1976 when it is applied to this series after it is standardized using the sample monthly means and variances for each of the 12 calendar months. Although Wang's (2003) method is not expected to perform well for all series with autocorrelated features, it worked well here anyway. This is attributed to an extremely significant mean-shift and minimal series autocorrelations.

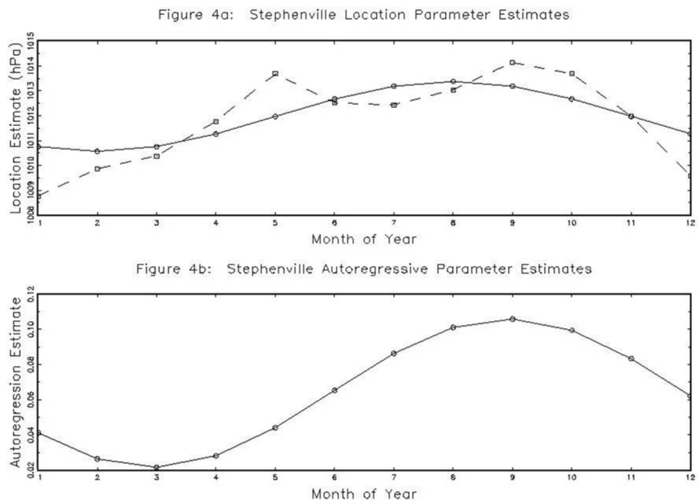


Figure 4. Estimates of the (a) location and (b) autocorrelation parameters for for the Stephenville pressure series. The dashed curve in (a) shows the sample averages of pressures in each month after adjusting for trend and the mean-shift.

To further illustrate the seasonality and autocorrelations in this series, Figure 4 plots estimates of m_v and f_v against the month v. In the plot of m_v in Figure 4a, sample averages of the month v pressures after adjusting for the trend and the changepoint mean-shift are also displayed (dashed curve). These values agree, to a rough order, with the three parameter cosine wave fitted in (10). One could explore adding a second order harmonic in this fit, but we will not do so here. The mean response of the fitted model is plotted against the data in Figure 1 and seems very reasonable. In

Figure 4b, it should be noted that the monthly autocorrelations range from 0.021 to 0.106, with the minimum occurring during March and the maximum in September. This amount of autocorrelation is not very heavy.

The methods have performed well for the Stephenville series, but, as mentioned, the changepoint was extremely obvious. We now move to a more difficult case, and one that will illustrate the full power of the methods.

b. A monthly temperature series

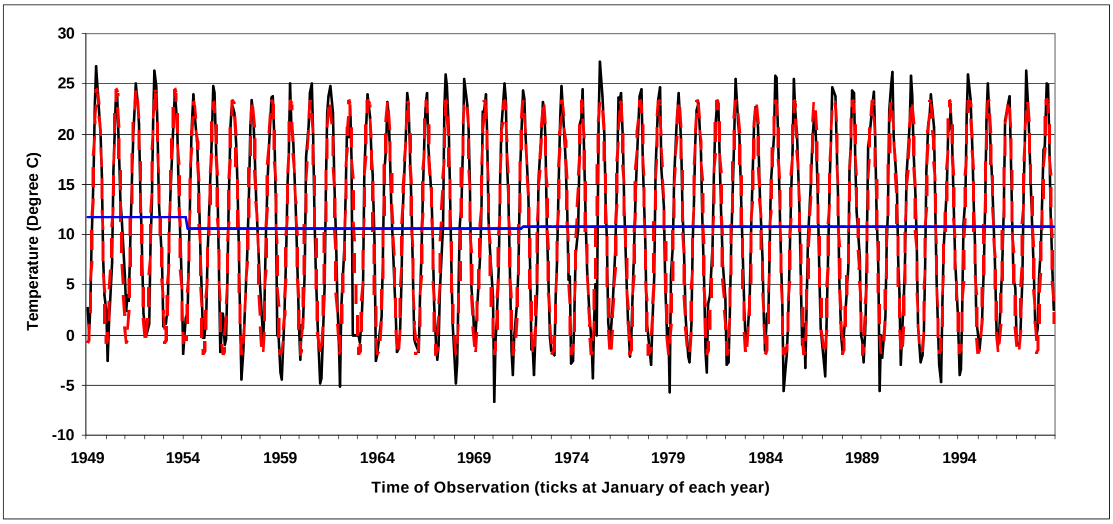


Figure 5. The same as in Figure 1 but for monthly averages of daily maximum temperatures recorded at Collegeville (Nova Scotia, Canada).

Figure 5 displays the series of monthly averages of daily maximum temperatures recorded at Collegeville (Nova Scotia, Canada) from 1949-1998 (50 years; $N = 600$). The seasonal cycle in the data is clear, with winter temperatures being colder and more variable than summer temperatures. With the seasonal structure of the time series viscerally dominating, it is hard to `eyeball' any changepoint here. However, as we show below, there is indeed some evidence for a changepoint. Versions of (6) and (7) were fitted to the series first without the scaling s_v^2 factors in the denominator. The F_c-statistics for this procedure peak at $c = 62$. This value of c was used to develop estimates of s_v^2, which are plotted in Figure 6.

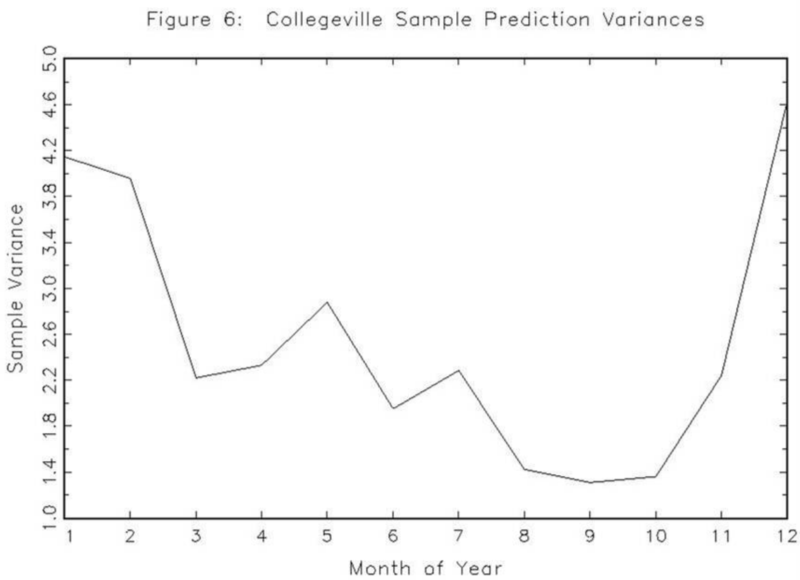


Figure 6. Sample prediction variances of the Collegeville temperature series.

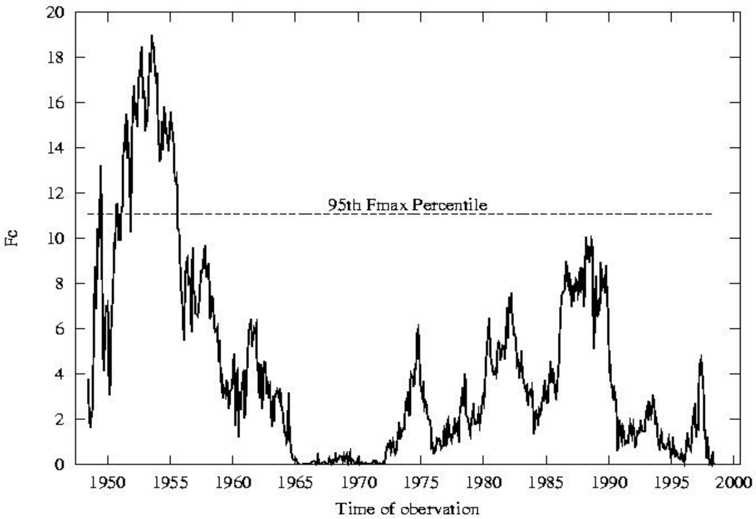


Figure 7. The F_c statistics of the Collegeville temperature series.

Figure 7 plots the F_c-statistics for this data weighting for the estimated values of s_v^2. The largest F_c statistic is $F_{\max} = 18.988$ and again occurs at c = 62 (February 1954) and exceeds the 95 percent confidence threshold. Hence, there is statistical evidence for a changepoint between February and March 1954. According to Vincent (1998), there was "a change in observer at the beginning of the 1950's along with a site relocation of about 10 km north of the previous site" and "the inspector subsequently reported that the new data had little in common with data from the former sites." However, the exact date of the 1950's relocation and observer change is not documented.

The other parameters of the full model fit with c = 62 are $\hat{\Delta} = -1.16$ºC, $\hat{b} = 0.000281$ºC/month, $\hat{A}_0 = 11.722$, $\hat{A}_1 = 12.960$, $\hat{t} = 7.398$, $\hat{B}_0 = 0.136$, $\hat{B}_1 = 0.128$, and $\hat{h} = 0.539$. The mean function of this model is plotted in bold against the data in Figure 5 and fits the series very well. Applying the multiple regression method to the same temperature series for 1916-95 with a reference series, Vincent (1998) also identified a changepoint between 1951 and 1952 (and another between 1935 and 1936), and hence added -1.6ºC to the annual means for 1936-51 (i.e., a decreasing step of 1.6ºC between 1951 and 1952) to homogenize the series. However, Vincent's estimates of step-size are based on the annual series (i.e., one datum per year) of length N = 60 and hence are more prone to sampling variability (note that N = 600 in our analysis). In addition, the model used to obtain the estimates does not include a linear trend component; ignoring a positive linear trend would lead to overestimation of the step-size. This is probably why Vincent's (1998) $\hat{\Delta}$ exceeds ours in absolute terms; in addition, the different periods of data used could have alsocontributed to the difference in the step-size estimate, and so couldthe existence of the documented changepoints in the early decades:
According to the station inspection reports, this station was relocated four times (in December 1926, August 1932, March 1936, and October 1948). Inclusion of these documented changepoints would complicate the analysis; this is why the period from 1949-1998 was selected for our analysis here. The two-phase regression approach of Wang (2003) was also applied to the same time series but standardized using the sample monthly means andvariances. This procedure identifies a changepoint between January and February 1963 to be of about 5% significance ($F_{\max} = 11.688$); it did not find the changepoint in the beginning of 1950's.

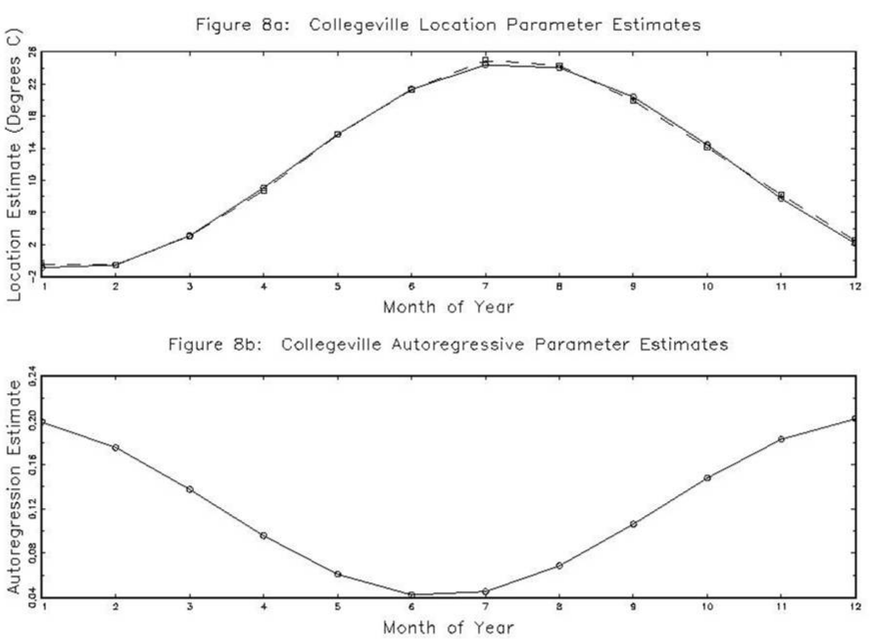


Figure 8. The same as in Figure 4 but for the Collegeville temperature series.

Figure 8 plots the model's estimated m_v and f_v for each month v. Notice how well the fitted sinusoid agrees with the empirically averaged estimated location parameters. One also sees substantially higher correlation levels in this series than for the Stephenville pressure series, with some of the f_v exceeding 0.20, further underscoring the danger in using methods which assume IID error structure.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. J. Climatol., 6, 661-675.

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity tests for linear trends. *International Journal of Climatology*, 17, 25-34.

Brockwell, R., and R. A. Davis, 1991: *Time Series: Theory and Methods*, 2nd Edition. Springer-Verlag, New York.

Brown, R. L., J. Durbin, and J. M. Evans, 1975: Techniques for testing the constancy of regression relationships over time (with discussion). *Journal of the Royal Statistical Society*, Series B, 37, 149-192.

Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate. *Journal of the Royal Statistical Society*, Series C, 53, 405-425.

Chen, J., and A. K. Gupta, 2000: Parametric Statistical Change Point Analysis. Birkhäuser, Boston.

Ducre-Robataille, J., L. Vincent, and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, 23, 1087-1101.

Easterling, D. R., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, 15, 369-377.
Gullet, D. W., L. Vincent, and L. Malone, 1991: Homogeneity testing of monthly temperature series: Application of multi-phase regression models with mathematical changepoints. Atmospheric Environment Service, Downsview Ontario, 47 pp.
Hanesiak, J. M, and X. L. Wang, 2005: Adverse weather trends in the Canadian Arctic. *J. Climate*, 18 (No. 16), 3140-3156.
Hawkins, D. M., 1977: Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72, 180-186.
Hinkley, D. V., 1969: Inference about the intersection in two-phase regression. *Biometrika*, 56, 495-504.
Hinkley, D. V., 1971: Inference in two-phase regression. *Journal of the American Statistical Association*, 66, 736-743.
Karl, T. R., and C. N. Williams, Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology*, 26, 1744-1763.
Kander, Z. and S. Zacks, 1966: Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Annals of Mathematical Statistics*, 37, 1196-1210.
Lund, R. B., H. Hurd, P. Bloomfield, and R. L. Smith, 1995: Climatological time series with periodic autocorrelations. *Journal of Climate*, 8, 2787-2809.
Lund, R. B., and J. Reeves, 2002: Detection of undocumented changepoints --- a revision of the two-phase regression model. *Journal of Climate*, 17, 2547-2554.
Lund, R. B., Q. Shao, and I. V. Basawa, 2006: Parsimonious Periodic Time Series Modeling. *Australian & New Zealand Journal of Statistics*, 48, 33-47.
Percival, D. B., and D. A. Rothrock, 2005: "Eyeballing" trends in climate time series: A cautionary note. *Journal of Climate*, 18, 886-891.
Page, E. S., 1955: A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42, 523-527.
Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review*, 109, 2040-2045.
Reeves, J., J. Chen, X. L. Wang, R. B. Lund, and Q. Lu, 2006: A review and comparison of changepoint detection techniques for climate data. Journal of Applied Meteorology and Climatology, accepted.
Rhoades, D. A., and M. J. Salinger, 1993: Adjustment of temperature and rainfall records for site changes. *International Journal of Climatology*, 13, 899-913.
Solow, A. R., 1987: Testing for climate change: an application of the two-phase regression model. *Journal of Climate and Applied Meteorology*, 26, 1401-1405.
Thompson, C. S., 1984: Homogeneity analysis of rainfall series: An application of the use of a realistic rainfall model. *Journal of Climatology*, 4, 609-619.
Vincent, L., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11, 1094-1104.
Wan, H., X. L. Wang, and V. R. Swail, 2006: A Quality Assurance System for Canadian Pressure Data. Journal of Applied Meteorology and Climatology, submitted (also in this report).
Wang, X. L., 2006: Climatology and Trends in some adverse and fair weather conditions in Canada, 1953-2004. JGR-Atmospheres, 111, D09105, doi:10.1029/2005JD006155.
Wang, X. L. and Y. Feng, cited 2004: RHTest user manual. Available at http://cccma.seos.uvic.ca/ETCCMDI/RHTest/RHTestUserManual.doc.
Wang, X. L., 2003: Comments on "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model". *Journal of Climate*, 16, 3383-3385.

ON HOMOGENIZATION OF CLIMATE LONG-TERM SERIES IN BULGARIA

Tania Marinova and Vesselin Alexandrov
National Institute of Meteorology and Hydrology,
66 Tsarigradsko chaussee, 1784 Sofia, Bulgaria
Tel. +359 2 9753986, Fax: +359 2 9884494
Tania.Marinova@meteo.bg, Vesselin.Alexandrov@meteo.bg

1. INTRODUCTION

Homogenization has become one of the basic elements of climatological studies (e.g. *Szalai, 2000*). An investigation of climatic change must be based on a homogeneous climatological time series (e.g. *Štepánek et al., 2000*). A series is said to be homogeneous "if its variations are caused only by variations in weather and climate" (e.g. *Conrad and Pollak, 1962*).

Inhomogeneities in climate time series arise from non-climatic factors like changes in station location, changes in methods to calculate means, changes in observation practices, changes in instruments and in station environment. Each of these changes may require a separate homogenization strategy. The changes may cause stepwise and/or gradual biases in the climatological time series, making these series unrepresentative of the climate of the concerning area (e.g. *Brandsma, 2000*).

Beside the well-known use in the climate change studies, more and more users request long-term time series in homogenized form. The overall trend shows a decrease of human observations, and a growing rate of automatization. In consequence, we often do not measure the same meteorological parameter, as earlier, only something similar to that, certainly new methods of observation are used, which imply rather different data quality problems, etc. The merging of satellite and radar information into the classical database could effect large breaks as well (e.g. *Szalai, 2000*).

The identification of local, regional and global climate change has become an important issue in climatology. Data homogeneity is strongly related to the climate change problem, which is at the centre of scientific and policy debates. It has been recognized and widely accepted that long and reliable observation series are required to address climate change issues and impact studies. Unfortunately, these high quality meteorological data series seldom exist, therefore it is imperative that homogenized data be used for theoretical and applied research (e.g. *Mersich, 1999*). As often clearly stated by the Intergovernmental Panel on Climate Change (IPCC), there is an urgent and continuing requirement for high quality and consistently collected observation and related homogeneous data sets to understand climate change, verify assessments and models use to generate future climate scenarios (e.g. *Scholefield, 1999*).

It was already mentioned that the long-term climatological time series are often plagued with discontinuities caused by station relocation, installation of new instruments, etc. Several types of disturbances can distort or even hide the climatic signal. Therefore, it is quite natural that the data are tested in order to locate possible discontinuities. However, usually the detection of the homogeneity breaks is not enough. The breaks appear to be so common that rejection of inhomogeneous series simply leave too few and too short series for further analysis. The widely adopted practice is to make adjustments in the non-homogeneous climatological time series (e.g. *Tuomenvirta, 1999*).

There are several direct and indirect methodologies for homogeneity testing. The direct methodologies include, for example, use of metadata, side by side comparisons of instruments, and statistical studies of instrument changes. The indirect methodologies consider use of single station data, development of reference time series, subjective and objective methods. The available objective methods include: Potter's method; Standard normal homogeneity test; two-phase regression; rank order change point test; Craddock test; Caussinus-Mestre technique; multiple analysis of series for homogenization (e.g. *Alexandersson and Moberg, 1997; Peterson et al., 1998; Szentimrey, 1999*).

Since 1994, Météo-France has put significant efforts on search, data rescue and homogenization of long series of weather measurement. These efforts allowed to built-up a base of homogenized data (e.g. *Moisselin and Mestre, 2002*). In Meteo-France, Toulouse, November, 2002 the homogenization method developed by Caussinus and Mestre (e.g. *Caussinus and Mestre, 1997; Mestre, 2000; Moisselin and Mestre, 2002*) was applied on climate long-term series from Bulgaria including data of average air temperature and precipitation (*Alexandrov et al., 2004*).

One of the major goals of the second study held in Meteo-France, Toulouse, from 17 November to 19 December, 2003 was to apply the French homogenization method on long-term series of sunshine duration from Bulgaria. Specific objectives were: to control monthly data of sunshine duration from selected climatological stations in Bulgaria; to detect breaks and outliers within the collected and controlled time series; to correct the currently used climate long-term series according to the defined breaks and outliers in order to obtain homogenized climate series; to validate the respective breaks.

2. LOCATION, EXPERIMENTAL MATERIAL AND METHODS

2.1 Location

Bulgaria (*Fig.1*) is located on the Balkan Peninsula in Southeastern Europe. The country includes 31% lowlands (0–200 m), 41% hills (200–600 m), 25% highlands (600–1600 m), and 3% mountains (> 1600 m). The Balkan Mountains split the country into Northern and Southern Bulgaria, and have a strong effect on the temperature regime. The country belongs to the temperate climate zone with a typical rotation of four seasons and variable weather throughout the whole year. Climate is continental to the north and close to Mediterranean to the south. The annual mean air temperatures in Bulgaria vary from –3.0 to 14.0°C, depending on the location and elevation. Air temperature normally reaches a minimum in January, and a maximum in July. The monthly mean temperature varies from –0.9 to 3.2°C in January and from 5.0 to 25.0°C in July. Total precipitation depends on the circulation patterns, site elevation, and the specificity of local orographic features. Annual mean total precipitation is approximately 500–650 mm, with an annual variation ranging from 440 to 1020 mm. The annual values of sunshine duration in the country are between 1800 and 2300 hours.

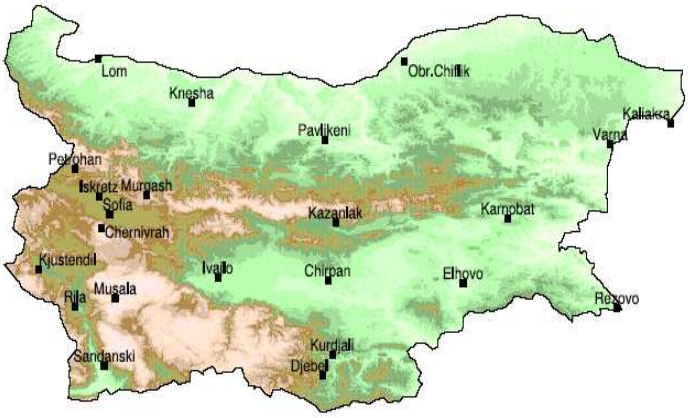


Fig.1. Climatological stations in Bulgaria with long-term records of sunshine duration, used in the study

2.2 Experimental material

Monthly data of sunshine duration from 22 Bulgarian climatological stations with long-term series were collected for the study (*Fig.1*, *Table 1*). All sunshine duration data applied in this study were provided by the Meteorological Database of the National Institute of Meteorology and Hydrology in Sofia, Bulgaria.

2.3 Methods

2.3.1. Data homogenization

The Caussinus-Mestre method, applied within this study, simultaneously accounts for the detection of unknown number of multiple breaks and generating reference series. It is based on the premise that between two breaks, a time series is homogeneous and these homogeneous sections can be used as reference series. Each single series is compared with others with the same climatic area by making series of differences (e.g. for minimum and maximum air temperature) or ratios (e.g. for sunshine duration). These differences or ratios series are tested for discontinuities. When a detected break remains constant throughout the set of comparisons of a candidate station with its neighbours, the break is attributed to the candidate station time series (e.g. *Caussinus and Mestre, 1997; Mestre, 1999, 2000; Peterson et al., 1998*).

Table 1. Climatological stations and series length of sunshine duration

Stations	**Beginning year (2001 – end)**	**Stations**	**Beginning year (2001 – end)**
Chirpan	1929	Varna	1961
Elhovo	1954	Kaliakra	1954
Karnobat	1931	Rila	1926
Kazanlak	1903	Cherni vrah	1936
Kjustendil	1968	Djebel	1937
Kneja	1942	Iskretz	1938
Kurdjali	1930	Ivajlo	1961
Lom	1954	Murgash	1954
Obrazcov chiflik	1903	Petrohan	1954
Pavlikeni	1933	Rezovo	1958
Sandanski	1951	Sofia	1954

2.3.1.1. Detection of breaks and outliers

For detection purposes, the formulation described by Caussinus and Lyazrhi (*1997*) is used. It allows the determination in a normal linear model of an unknown number of breaks and outliers. They formulated it is a problem of testing multiple hypotheses. Let us give now the formulation of this procedure in the case of a normal sample (e.g. *Moisselin and Mestre, 2002*). We consider n normal random variables Y_i ($i=1,...,n$) and let Y denote the column vector of the Y_i's. We assume that the probability distribution of Y is n-dimensional normal, with covariance matrix I_n (identity matrix of order $n \times n$) up to the unknown variance σ^2.

Let k be the number of breaks and l – the number of outliers. Let $\tau_1, \tau_2, ..., \tau_k$ be the positions of the k breaks, and let $\delta_1, \delta_2, ..., \delta_l$ be the positions of the l outliers. Let $K = (\{\tau_1, \tau_2, ..., \tau_k\}, \{\delta_1, \delta_2, ..., \delta_l\})$ be the set of breaks and outliers. To simplify the notation, we will set $\tau_0=0$ and $\tau_{k+1}=n$. Finally, let $\Delta = \{\delta_1, \delta_2, ..., \delta_l\}$ and $n_j = \tau_j - \tau_{j-1} - Card[\{\tau_{j-1}+1, \tau_{j-1}+2, ..., \tau_j\} \cap \Delta]$, i.e. n_j is equal to the length of the period [$\tau_{j-1}+1, \tau_j$] minus the number of outliers within this period.

We denote:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \ , \qquad \bar{Y}_j = \frac{1}{n_j}\sum_{i=\tau_{j-1}+1}^{\tau_j} Y_i \ , \qquad j=1,\ldots,k+1;\ i \notin \Delta \qquad (1)$$

Let:

$$C_{\varnothing}(Y)=0$$

$$C_K(Y) = ln\left[1 - \frac{\sum_{j=1}^{k+1} n_j \left(\bar{Y}_j - \bar{Y}\right)^2}{\sum_{i=1}^{k+1}\left(\bar{Y}_i - \bar{Y}\right)^2}\right] + \frac{2(k+l)}{n-1} ln(n) \qquad (2)$$

The penalized log-likelihood procedure proposed by Caussinus and Lyazrhi (1997) is:

$$select\ H_{K^*}\ such\ that\ K^*=Argmin_K(C_K(Y)) \qquad (3)$$

The variance σ^2 is estimated by

$$\frac{1}{n-k-l-1}\sum_{j=1}^{k+1}\sum_{i=\tau_{j-1}+1}^{\tau_j}\left(Y - \bar{Y}_j\right)^2 ,$$

where the number and positions of outliers and breaks are those given by (3).
The procedure (3) has been proved to be asymptotically Bayes invariant optimal under a set of assumptions, which turn out to be realistic (e.g *Mestre, 2000*) in the problem we are dealing with. For the particular problem of breaks in a Gaussian sample, the chosen penalty term gives much better results than Akaike's or Schwartz's criteria (e.g. *Moisselin and Mestre, 2002*).

The natural way to compute the procedure is to calculate $C_K(Y)$ for every possible hypothesis H_K (complete procedure). Nevertheless, this approach suffers from a major drawback: the number of hypotheses to examine rises very fast with n (length of the series) and $k+l$ the number of accidents to be detected. When detection is only performed for breaks, a dynamic programming algorithm (e.g. *Hawkins, 2001; Lavielle, 1998*) can be used. The computation time then becomes only linear in k, and quadratic in n. To enable the detection of outliers at a reasonable computing cost, a slightly different algorithm (*Mestre, 2000*) is used.

At each step, one or two more breaks are added to the previous selected hypothesis. Analytical studies (e.g. *Mestre, 1999*) show that this double step procedure gives better detection results than the single step procedure for up-and-down breaks (and without significant improvement for staircase configuration). Furthermore, a triple step procedure, much more greedy in terms of computation time, leads to small improvements (e.g. *Mestre, 1999, 2000*). The Causinus-Mestre method, with a double step procedure, is now the standard detection part of the homogenization method used in Météo-France (e.g. *Moisselin and Mestre, 2002; Moisselin et al., 2002*).

2.3.1.2. Correction of breaks and outliers

The knowledge of break positions can be a very interesting aspect for some users. For many applications (such as climate change studies) it is the first half-part of the problem. The other one, described below, is the break correction.

A two factors linear model is proposed for correction purposes (e.g. *Mestre, 2000*). The series within the same climatic area are considered to be affected by the same climatic signal factor at each time, while the station factor remains constant between two breaks. The model is applied after break detection. It provides the correction coefficient of a set of inhomogeneous series, through weighted least-squares estimation of the parameters. The weighted least squares allow correction of series with missing data. It also allows the data weighting, according to their supposed quality, which can be estimated, for example, with the correlation between the stations.

The above formulation is equivalent to an exact modelling of the relative homogeneity principle. Given a set of inhomogeneous instrumental series, it allows unbiased estimations of the breaks affecting these series. This method does not require computation of regional reference series, and is currently the standard correction part of the homogenization method used at Météo-France (e.g. *Moisselin and Mestre, 2002; Moisselin et al., 2002*).

3. RESULTS AND DISCUSSION

3.1 Data homogenization – control, break detection and correction

The homogenization process was performed on sets of 13 series of sunshine duration respectively, merged with geographical criteria. The first step was the performance of a quality control of the long-term series of the weather elements used in the study. The control procedure was executed several times till appropriate data sets were obtained. The anomalies of monthly sunshine duration for each year and station were compared and analyzed in order to locate and remove possible data errors (*Fig.2*). The obvious crude errors as well as some suspicious values of sunshine duration were reported to the Division of Climatology and Weather Network as well as to the Meteorological Database Management Division. The updated data were again checked out by the controlling software. The remained errors and suspicious values this time were replaced by the respective value for missing data (i.e. –999.9).

The second step in the homogenization procedure was to replace missing monthly values assuming that these values are very few and their replacement would not have any impact to the data series. The two factors linear model by means of the computed weighted least squares allows correction of series with missing data. For this purpose, the linear model was run with the option for correction of missing data.

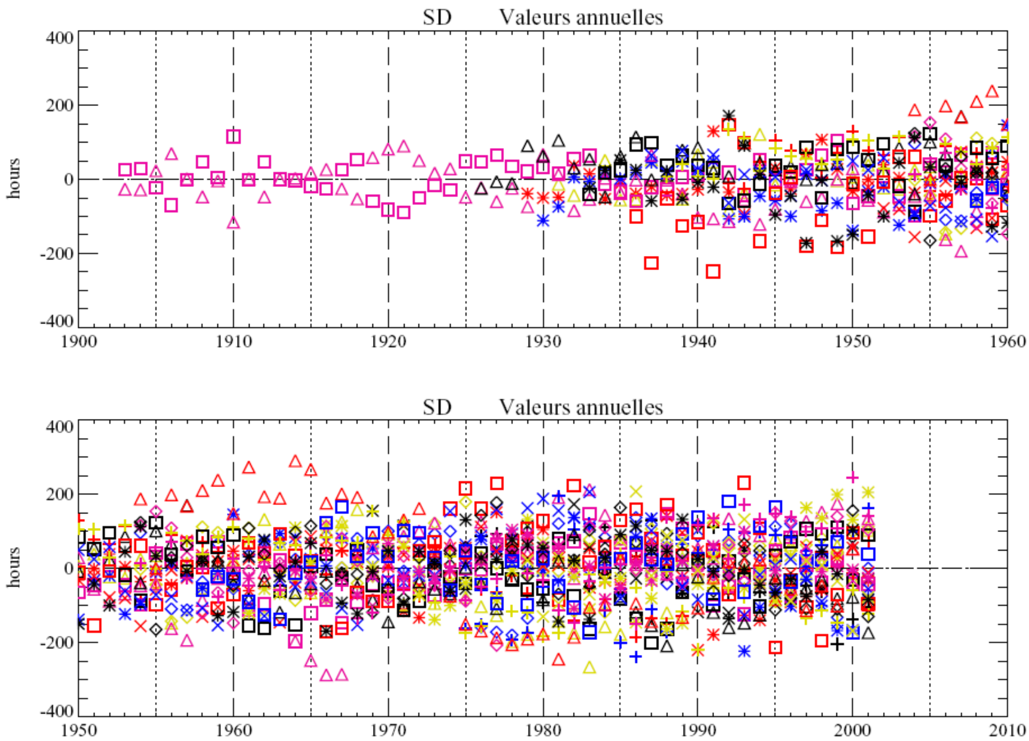


Fig. 2. Anomalies of annual sunshine duration sums for all climatological stations, used in the study

The next step was to calculate the respective ratios for sunshine duration. These ratios were then tested to put into evidence breaks or outliers. The typical homogenization techniques are based on the assumption that climatic variations affect in the same way a homogeneous regional reference series, whose reliability cannot be proved. The different methods (e.g. *Alexandersson, 1986; Førland and Hanssen-Bauer, 1994; Peterson and Easterling, 1994*) for creating such series do not guarantee their perfect homogeneity.

There is an easy way to get round the reference series. It is based on the simple statement that between two breaks a series is reliable (by definition), so these sections can be used as reference series (e.g. *Mestre, 2000*). Each single series is compared to others within the same climatic area by making a series of ratios. These series are then tested for discontinuities.

At this stage, it is not known which individual series is the cause of a shift detected on a ratio or series. However, it was already mentioned that according to the Caussinus-Mestre method, if a detected break remains constant throughout the set of comparisons of a candidate station in respect to its neighbours, it can be attributed to this candidate station. The detection of the outliers follows the same principle.

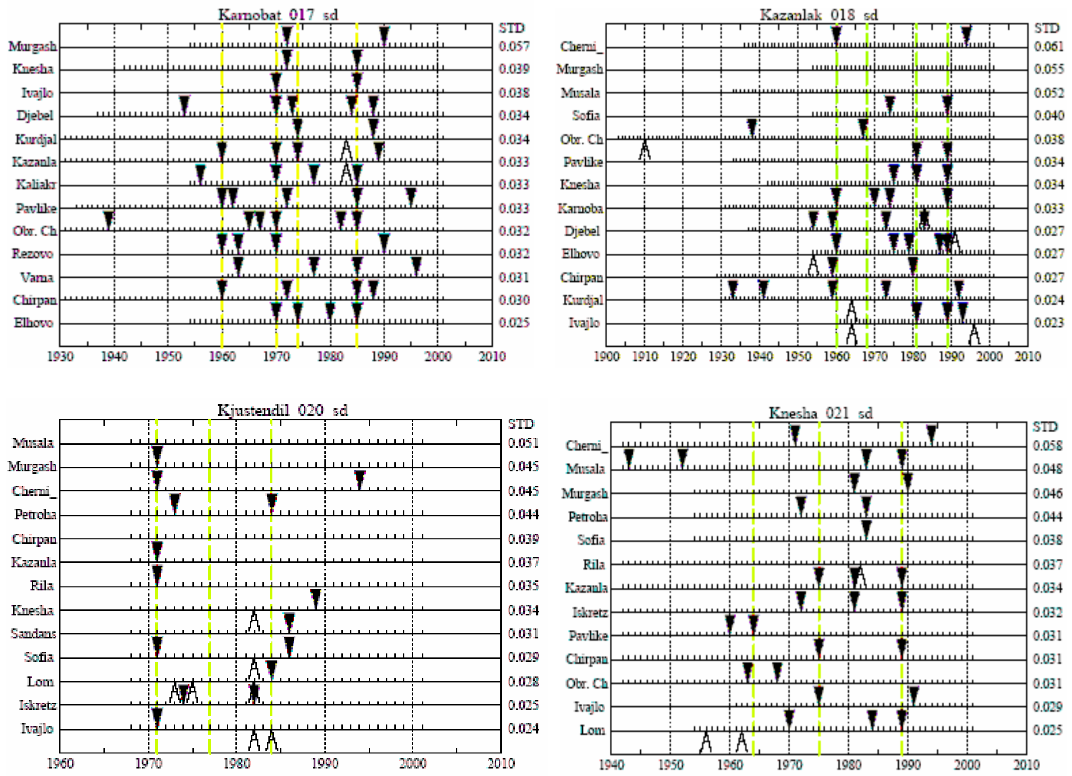


Fig. 3. Homogenization of sunshine duration data in 4 climatological stations:

▼ – break; A – outlier; dash line – validated break and then corrected

Ratio series were computed and constituted between all climatological stations, used in the study, and their respective neighbour climatological stations. The breaks and outliers were then put into evidence by the double-step procedure applied within the Caussinus-Mestre method. For example, some detected breaks and outliers of sunshine duration are shown in *Fig.3*. The black triangles indicate the position of the detected breaks in the ratio series of the presented climatological station versus the other climatological stations, while A points out the outliers. The climatological stations are ordered from the top to the bottom with respect to increasing values of the estimated standard deviation STD. Hence, in practice, the reliability of the comparisons slightly decreases from the bottom to the top.

Several breaks during the 20th century can be detected easily in *Fig.3*, considering the relatively good alignment of breaks in sunshine duration. For example, in climatological stations Karnobat, Kazanlak and Kjustendil the breaks of sunshine duration data in 1970, 1985 (Karnobat), 1989 (Kazanlak) and 1971 (Kiustendil) respectively, are obvious.

The knowledge of break positions for many applications including climate variability and change studies is the first important half of the final goal. The second part of the homogenization goal is the break correction. The two factors linear model was applied after break detection and validation. It was assumed that the series within the same climatic area are considered to be affected by the same climatic signal factor at each time, while the station factor remains constant between two breaks. The model computed the correction coefficients of a set of non-homogeneous series, through weighted least squares estimation of the parameters.

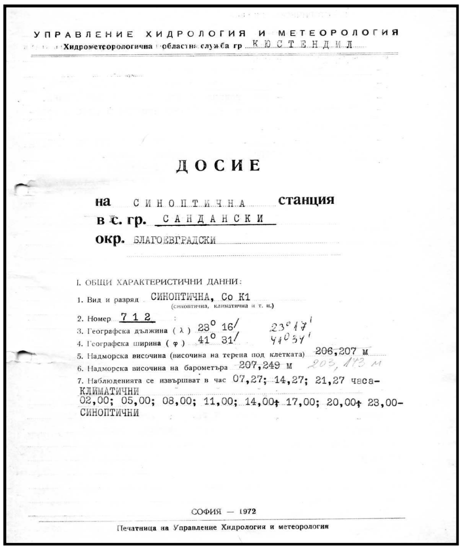
It was impossible to locate straightaway all possible breaks: the pronounced breaks hided smaller one's. Thus, the procedure of detection and correction of breaks and outliers was not automatic. It was iterative and the expert knowledge and strategy was very essential. Every time the expert team validated the breaks keeping in mind some statistical and climatological issues. The whole procedure of break detection and correction took time – it was near 15 times in order to locate, validate and correct all the breaks and outliers in the series of sunshine duration. The whole iteration of homogenization of sunshine duration ended when all or most break risk was gone (*Fig.3*).

4. LIMITATION

The major limitation was the lack of metadata at the time the study was implemented. Although the Meteorological Database Management Division at the National Institute of Meteorology and Hydrology in Sofia, Bulgaria has initiated a work on this problem digitized metadata are not fully available yet.

It is clear that the Bulgarian weather data are influenced by a wide variety of parameters like the environment, the instrumentation, observing practices, data processing and others. This means that for each single data we should know where and how the measurement was made. For a historical long-term climate time series this knowledge would lead to a complete station history. Unfortunately, our knowledge of station history most likely will not be 100% complete, nevertheless greatest efforts should be undertaken to study metadata. Metadata should be treated with the same care as the data themselves (e.g. *Auer, 2003*).

For all synoptic, climatological and precipitation stations, from the National hydrometeorological network on the territory of the country, there are paper records including description of the station and its environment as well as detailed information about station activities, since the beginning of the respective measurements till now. It turned out that there were some omissions in these documents. For this reason we work in two directions: (a) digitization of the available station documents in the Meteorological Database Management Division (*Fig.4*); (b) station documents update in the Regional Centres of the National Institute of Meteorology and Hydrolojy in Pleven, Varna, Plovdiv and Kjustendil.


УПРАВЛЕНИЕ ХИДРОЛОГИЯ И МЕТЕОРОЛОГИЯ
Хидрометеорологична областна служба гр КЮСТЕНДИЛ

ДОСИЕ

на СИНОПТИЧНА станция
в с. гр. САНДАНСКИ
окр. БЛАГОЕВГРАДСКИ

I. ОБЩИ ХАРАКТЕРИСТИЧНИ ДАННИ:
1. Вид и разряд СИНОПТИЧНА, Со К1
(синоптична, климатична и т. н.)
2. Номер 7 1 2
3. Географска дължина (λ) 23° 16′ 23°17′
4. Географска ширина (φ) 41° 31′ 41°34′
5. Надморска височина (височина на терена под клетката) 206,207 м
6. Надморска височина на барометъра 207,249 м 203,143 м
7. Наблюденията се извършват в час 07,27; 14,27; 21,27 часа-КЛИМАТИЧНИ
02,00; 05,00; 08,00; 11,00; 14,00; 17,00; 20,00; 23,00-СИНОПТИЧНИ

СОФИЯ — 1972
Печатница на Управление Хидрология и метеорология

Fig. 4. The title page of a file from synoptic station Sandanski

5. CONCLUDING REMARKS

The need for reliable data becomes more and more apparent, both in space and time, because too much is at stake to rely on inaccurate data. The very existence of our society is threatened. Therefore, it is important for all WMO members (including Bulgaria) to produce and make available homogeneous series of data and corresponding metadata.

The results of this study show that homogenization is important for building of reliable meteorological database in Bulgaria. It is obvious that homogeneous weather series of data are essential for research. For producing high quality time series efficient measures for testing the homogeneity should be applied. The French homogenization procedure, which is applied in Météo-France, was proved in the study as an essential tool. By directly comparison of each climate long-term series to its neighbours, it was shown that problem with construction of homogeneous reference series does no longer exist. The applied methodology of homogenization is valuable for practical use such as on climate data in Bulgaria, even with missing metadata, and allows the detection of multiple breaks. Most homeogenization methods in Europe have been developed for the analysis of temperature and precipitation only. However, the Caussinus-Mestre method for the relative homogeneity testing of climatological series and the model performing correction of non-homogeneous climate series were also successfully tested on long-term series of sunshine duration. In fact, the executed homogenization was very useful for better understanding of sunshine duration series in Bulgaria.

One of the most important problems in the climate research is the quality of data. Long series of reliable climatological data are required in climatological studies on the natural climate variability and the effect of anthropogenic influences on recent climate. However, high quality climatological data seldom exist because in reality many types of disturbances can affect the respective climate series. Many efforts were put in this study for quality control on the Bulgarian series of sunshine duration. It should be stressed that the respective series were affected by different types of errors. Therefore, it is recommended that before any homogeneity testing of Bulgarian weather data to be applied, an extensive routine quality control has to be performed.

Historical time series carry the information of natural and artificial variability. However, before climate variability can be studied all artificial biases have to be removed. This is a hard job but unavoidable. Although this problem could be treated by using one or several homogenization techniques, metadata will provide a better insight and explain the reasons of breaks and support the statistical test results. It is always advisable to compare what station history says and what data analysis identifies (e.g. *Auer, 2003*). The importance of metadata was assumed by the WMO Commission for Climatology and the working group on Climate Change Detection (e.g. *Niedzwiedz and Ustrnul, 2000*). A proposal was given for Global Climate Observing System Surface Network Sites (e.g. *WMO, 1999*). Also WMO has stressed its strong interest in metadata recording for current measurements, but also in metadata recovery for historical time series. This interest is underlined by the establishment of a WMO Expert Team on Metadata for Climate Applications within the Commission for Climatology. The expert team members have been preparing guidelines on metadata and homogenization (e.g. *Auer, 2003*). These guidelines would be of help when respective metadata are gathered in Bulgaria.

Acknowledgements

Acknowledgements are addressed to the French Meteorological Society (SMF) and Météo-France for the support and opportunity to use facilities and respective software at the French Weather Service. Thanks also to the Direction of Climatology in Météo-France for the support to the study. Many thanks to Jean-Marc Moisselin and Brigitte Dubuisson (Direction of Climatology) for their useful help in order to apply the homogenization procedure on Bulgarian sunshine duration data.

REFERENCES

Alexandersson, H., 1986. A homogeneity test applied to precipitation data. Int J Climatol 6: 661–675.

Alexandersson H, Moberg A, 1997. Homogenization of Swedish temperature data. Part: A homogeneity test for linear trends. Int J Climatol 17:25–34.

Alexandrov, V., M. Schneider, E. Koleva and J-M. Moisselin, 2004. Climate Variability and Change in Bulgaria during the 20th Century. Theoretical and Applied Climatology 79(3–4): 133–149.

Auer, I., 2003. Metadata and their role in homogenization. Abstracts of the fourth seminar for homogenization and quality control in climatological databases, 6–10 October 2003, Budapest, Hungary.

Brandsma, T., 2000. Weather-type dependent homogenization of the daily Zwanenburg/De Bilt temperature series. Proceedings on the third seminar for homogenization and quality control in climatological databases, Budapest, Hungary 25–29 September 2000 (www.met.hu)

Caussinus, H. and O.Mestre, 1997. New mathematical tools and methodologies for relative homogeneity testing. Proceedings of the First Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 63–82.

Caussinus and F. Lyazrhi, 1997. Choosing a linear model with a random number of change-points and outliers. Ann. Inst. Statist. Math., 49, 761–775.

Conrad, V., L. W. Pollak, 1962. Methods in Climatology. Harvard University Press, Cambridge, Massachusetts, 459.

Førland, E.J., Hanssen-Bauer, I., 1994. Homogenizing long Norwegian precipitation series. Journal of Climate 7:1001–1013.

Hawkins, D.M., 2001. Fitting multiple change-points to data. Comput Statist Data Anal 37:323–341.

Lavielle, M., 1998. Optimal segmentation of random processes. IEEE Transactions on Signal Processing 46(5):365–1373.

Mersich, I., 1999. Open address. Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, WMO–TD, No.962, WMO, Geneva

Mestre, O., 1999. Step-by-step procedures for choosing a model with change points. Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, WMO–TD, No.962, WMO, Geneva, 15–26.

Mestre, O., 2000. Méthodes statistiques pour l'homogénéisation de longues séries climatiques. Thèse de doctorat de l'Université Paul Sabatier. Toulouse: Université Paul Sabatier

Moisselin, J.M. and O.Mestre, 2002. Research, Digitisation and Homogenization of Long-Term Data Series. Actes du Colloque "La Pérennisation et la Valorisation des données scientifiques et techniques" (Ensuring Long-Term Preservation and adding Value to scientific and technical data) organisé par le CNES du 5 au 7/11/202, Toulouse.

Moisselin, J.M., M. Schneider, C. Canellas and O.Mestre, 2002. Les changements climatiques en France au XXe siecle. Etude des longues séries homogénéisées de données de température et de précipitations. La Météorologie 38 : 45–56.

Niedzwiedz, T. and Z. Ustrnul, 2000. Proposal of the metadata base by using of the MS access database. Proceedings on the third seminar for homogenization and quality control in climatological databases, Budapest, Hungary 25–29 September 2000 (www.met.hu)

Peterson, D.C., DR Easterling, 1994. Creation of homogeneous composite climatological reference series. Int J Climatol 14:671–679.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullet D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey, Salinger J, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998) Homogeneity adjustments of *In Situ* atmospheric climate data: a review. Int J Climatol 18:1493–1517.

Scholefield, 1999. Open address. Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, WMO–TD, No.962, WMO, Geneva

Štepánek, P., R. Brázdil and V. Kveton, 2000. Homogenization on air temperature series in the Czech Republic. Proceedings on the third seminar for homogenization and quality control in climatological databases, Budapest, Hungary 25–29 September 2000 (www.met.hu)

Szalai, S. 2000. Preface. Proceedings on the third seminar for homogenization and quality control in climatological databases, Budapest, Hungary 25–29 September 2000 (www.met.hu)

Szentimrey, T., 1999. Multiple analysis of series for homogenization (MASH). Proceedings of the second seminar for homogenization of surface climatological data, WCDMP 41/WMO–TD 962. Geneva: WMO, 27–46.

Tuomenvirta, H., 1999. The influence of adjustments on climatological time series., Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, WMO–TD, No.962, WMO, Geneva

World Meteorological Organization (WMO), 1999. Global Climate Observing System (GCOS), GCOS Surface Network (GSN), Data Format and Supporting Documentation for WMO Members to Use When Providing digital Historical Data for GCOS Surface Network Sites to the National Climatic Data Center, P/GSN – Annex 4, 1–17.

HOMOGENIZATION OF A DENSE THERMO-PLUVIOMETRC MONTHLY DATABASE IN THE BALEARIC ISLANDS USING THE FREE CONTRIBUTED R PACKAGE “CLIMATOL”

José A. Guijarro
Instituto Nacional de Meteorología, CMT en Illes Balears, Palma de Mallorca
jaguijarro@inm.es

ABSTRACT

"Climatol" is a set of routines for climatological applications than run under the multi-platform statistical package "R", freely available at "http://cran.r-project.org/". It is mainly devoted to the homogenization of monthly series, though may be applied to daily data as well. The homogenization method is based on comparing each test series with a reference series constructed for the same station through interpolation of ratios, differences or standardized values of the surrounding stations. This method avoids the use of regression techniques, with the advantage of been more robust and simple, and, most importantly, enabling the use of data from nearby stations when there is no common period of observation. The comparison of the problem series with their estimated references allows the detection of point errors, shifts and trends through standard statistical tests, optionally showing graphical representations of the results. The computed reference values may be readily used to fill the missing data of the series. The application of these methods to a dense thermo-pluviometric monthly database in the Balearic Islands showed a wide variety of situations, indicating the convenience of using an iterative strategy, thereby detecting and correcting only the coarser errors in the first place, and leaving the less prominent ones to the following iterations.

1. INTRODUCTION

The problem of coping with inhomogeneities of the climatological series is as old as the series themselves, and has been addressed by a multitude of investigators that have applied a variety of methods to detect point errors, sudden changes in the averages and anomalous trends (see Peterson et al., 1998, for a review). Yet no definite methodology has been already established, because some may be more appropriate than others depending on the climatological variable studied, the climatic regime and physiographic complexity of the area, the density of the observing networks, and even the final purpose of the data set.

The detection of inhomogeneities relies on the comparison of the problem series with a reference one that should be homogeneous and well correlated with the former. The reference series may be that of the best (long and homogeneous) station of the same climatic area, but as it is quite difficult to find long series that have not suffered changes of location, instrumentation or local conditions, it is generally preferable to build the reference as a combination of series from several stations (Peterson and Easterling, 1994), using as weighting factor some function of their correlation, distance, or both.

Once the reference series has been obtained, it can be used to determine which variations in the problem series are due to the climate variability and which are real inhomogeneities that should be corrected. The latter may be supported by the history of the station, through registered dates of any change that might have affected the observations (metadata), but these are often incomplete, and sometimes totally absent.

This work presents an implementation of a simple method for the detection of inhomogeneities in a climatological database, and discuses the first conclusions of their application to the homogenization of thermo-pluviometric monthly data from the Balearic Islands (Western Mediterranean).

2. METHODOLOGY

Climatic data bases are usually formed by a few long and complete series mixed with many short records with random beginning and end dates, and often with interspersed missing data spells. Regression techniques are extensively used for estimating these missing data from well correlated stations, but this prevents the use of nearby stations when there is not a common period of observations or it is too short to derive reliable regression equations. This is a frequent situation when an observatory changes its location (e.g., from near an airport terminal to a runway head): old and new stations are the best reference for each other, but no regression can be established. (Even when a common observing period of a couple of years is maintained, that supposes only two common terms in each monthly series). On the other hand, to maximize the existence of common observational periods between stations, many homogeneity studies select only the longer series of a data base, thereby disregarding a lot of potentially valuable information from many shorter records.

Here the priority has been focused on taking advantage of all the available climatic information. To achieve this goal regression techniques have been substituted by the simpler method of Paulhus and Kohler (1952) that applies a spatial interpolation of rates to the normals to fill daily precipitation data. This method was compared by Young (1992) with those of optimum interpolation (Gandin, 1963) and multiple discriminant analysis, and was the only that produced unbiased estimations while suffering the lowest reduction of variance. Their RMS errors were slightly higher than those of the multiple discriminant analysis, but lower than the optimum interpolation ones. This are convenient properties, since unbiased estimations will produce the best normal values (e.g., for climatic maps), while the reduction of variance minimization is a necessity in studies of variability and extreme value probabilities.

Proportions to the normals seem appropriate for precipitation or wind speed for instance, but for temperatures (among other variables) are better to use standardized values. This implies to know the averages and standard deviations of all series for a common and long period of observation, an incompatible constraint with our fragmented data set. Therefore, averages (and standard deviations if needed) are computed firstly with the available data, and missing data are filled with the unstandardized reference series computed for each station. This allows recalculating the means of all series for the complete period of study. As the new averages will differ from the previous, the reference series must be recomputed, and this process is repeated until the maximum difference of averages is lower than a prescribed threshold.

As previously stated, a reference series is computed for each station. Paulhus and Kohler's original method used only the three nearest stations, and other authors also limit the number or reference stations or impose a maximum distance (e.g. Romero et al., 1998). Here all available data are used to compute each reference series, weighting them by a function of distance. This way, the method is flexible enough to adapt to the varied distributions of neighbouring stations than may happen in a database in different periods.

The chosen weighting function was a/(a+d^2) or, dividing by 'a': 1/(1+d^2/a), where 'a' is a shape parameter controlling the relative weight of nearby stations with respect to the more distant ones (Fig. 1).

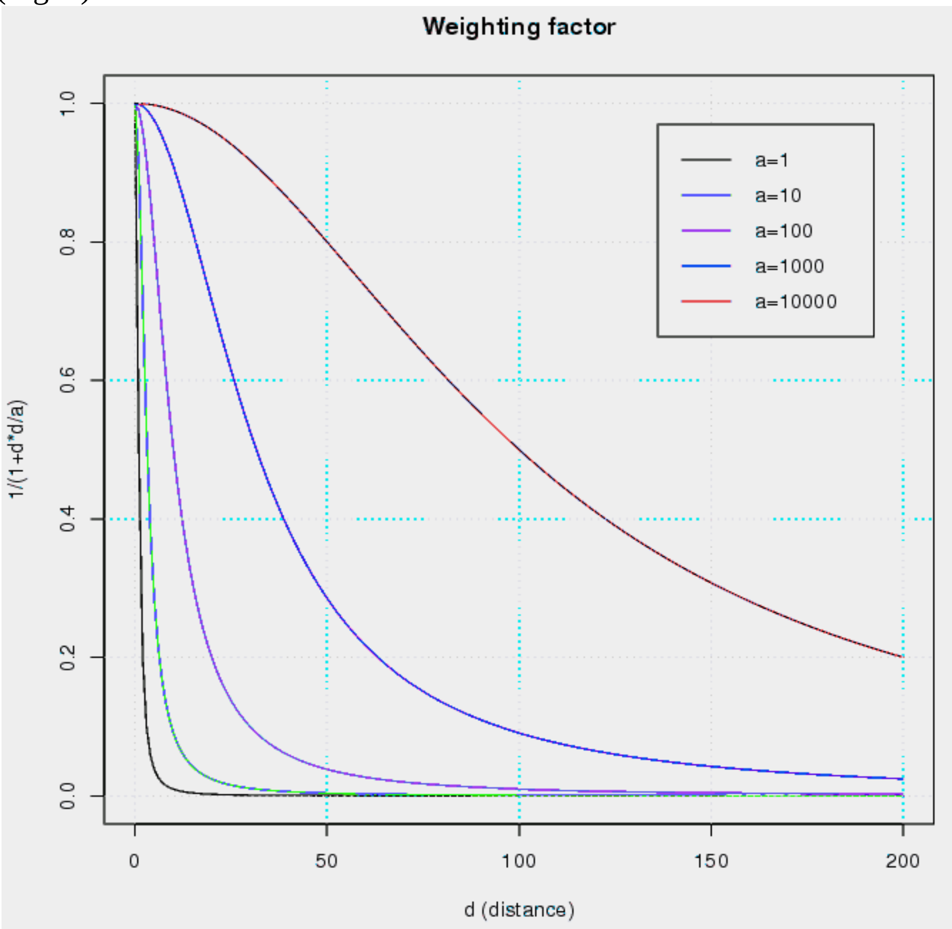


Fig. 1: Weighting factor (inverse to the square distance between stations) modulated by the shape parameter 'a'.

When the iterative process of computing the reference series for each station is completed, the following step is to compare, for each station, the observed and computed series. This is done with the ratios to normals or full standardized series (differences from normals is another option offered to the investigator), and a new series is calculated subtracting the reference from the observed standardized series. Differences (used e.g. by Aguilar et al., 1999) are preferred over quotients (as used by Alexandersson, 1986) because they can be applied to other variables than precipitation, and to rainfall in arid places, where monthly means near to zero cause problems in the computations of ratios (Almarza et al., 1994).

Where both series are homogeneous, the series of differences should behave as a random variable (white noise). In practice, three main types of inhomogeneities may be present: 1) Point errors (coming from observation to transmission and mechanization processes); 2) Shifts in the mean (changes of location, instrumentation, observing practices or land use of the surroundings); and 3) Trends (sensor decalibration, urban growth). And all of them may be present in real records (Fig. 2).

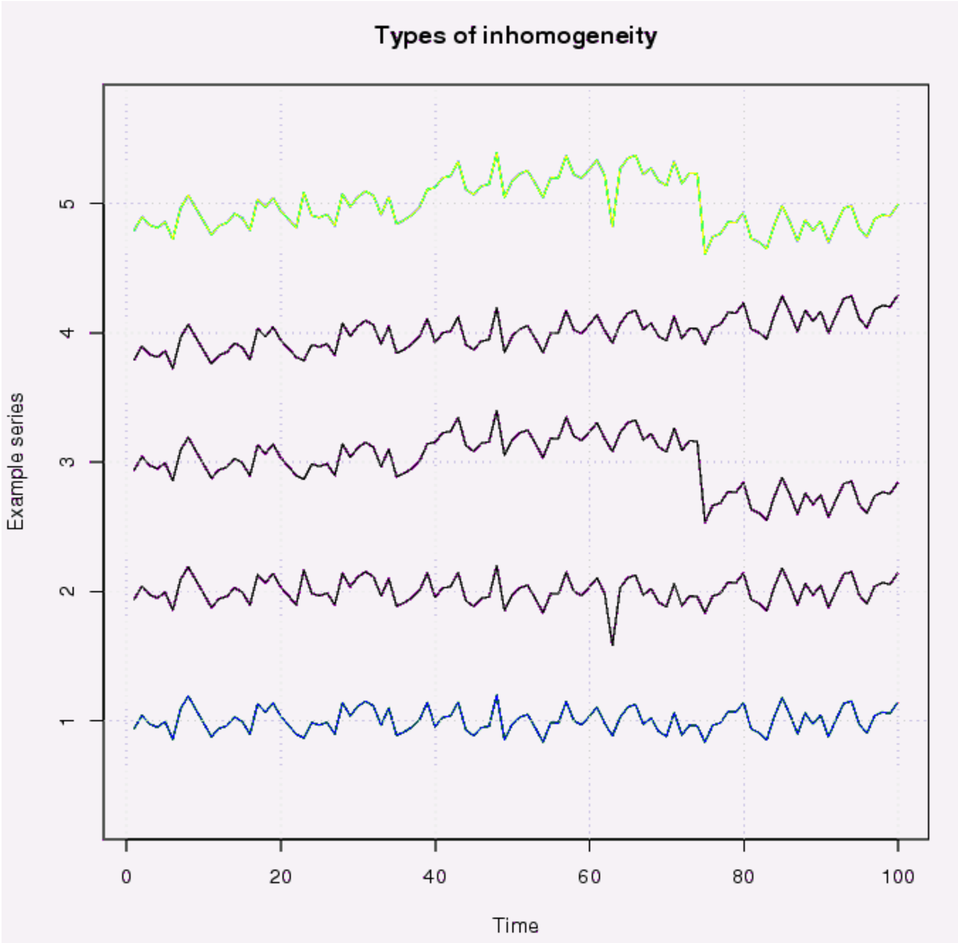


Fig. 2. Common inhomogeneities in the difference series: 1) Control, homogeneous (random noise). 2) Two point errors of +/- 3 s.d. (standard deviations); the first, at term 22, unnoticeable. 3) Two shifts in the mean of +2 and -3 s.d. 4) Trend of -1.5 to +1.5 s.d. 5) All previous inhomogeneities together.

Graphic representation of this series allows for visual inspection, and p-values of possible shifts and trends are computed from running t-tests (on 10 and 20 terms moving windows) and regression with time.

This methodology has been implemented as a contributed package to the statistical system "R", which is free, multi-platform (there are versions for different computer architectures), and runs under different operating systems, thus allowing its use in a wide variety of working environments. Moreover, since beeing open software, investigators may modify its routines to adapt them to their particular needs and contribute to their improvement.

3. APPLICATION TO A DENSE THERMO-PLUVIOMETRIC DATABASE

The Balearic islands lie in the Western Mediterranean, at 100 km of mainland Spain. Their 5000 km^2 are distributed in four major islands and many smaller isles. They have a varied orography, with mountain ranges with summits up to 1440 m (in Majorca), plains, and undulated terrains. The climate is typically Mediterranean, with a dry summer and a temperate winter. The maximum monthly average precipitation occurs in October, with 15% of annual precipitation, and the minimum corresponds to July, with means normally under 10 mm and medians of around 2 mm.

As most of the precipitations are due to convective processes, they have a rather spotty distribution. To catch this spatial variability, the pluviometric network was increased in the sixties to around 170 rain gauges, based mainly on amateur cooperators (figures 3 and 4).

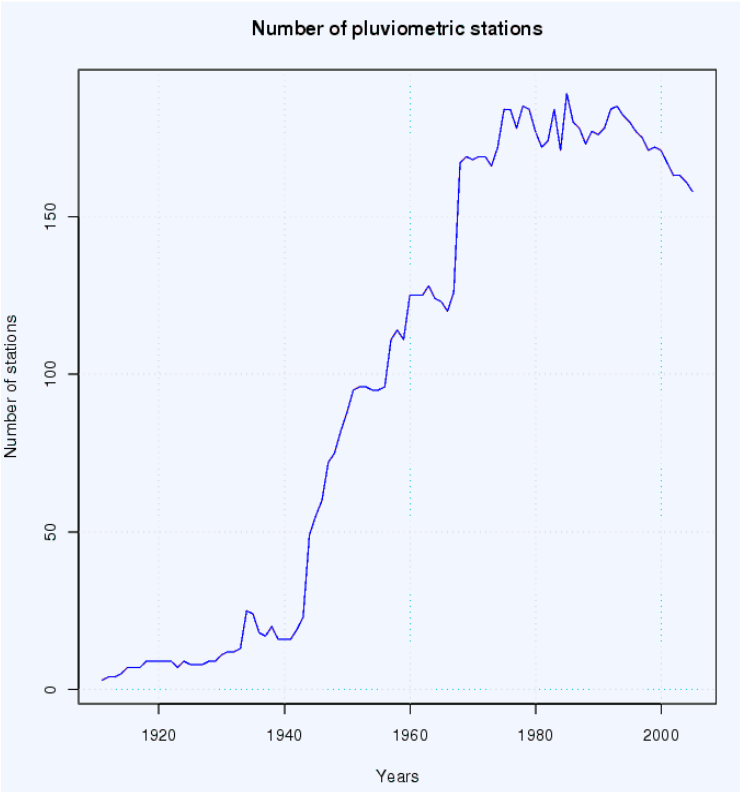


Fig. 3. Number of rain gauges in the Balearic Islands along time

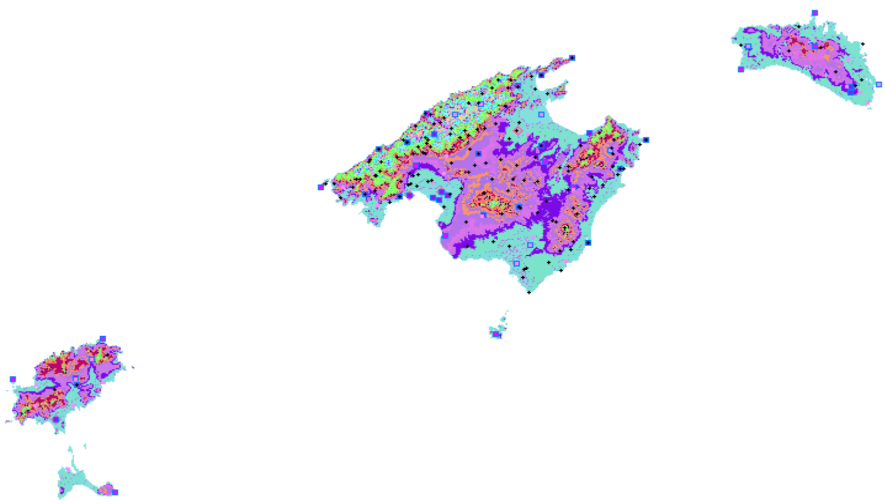


Fig. 4. Observing network in the Balearic Islands. Black crosses measure only daily precipitation; the rest measure at least daily precipitation and temperature extremes. (Majorca, the major island, is about 100 km wide).

For the application of the "climatol" methodology to the monthly series of this network, a selection was made of all series that had a minimum period of observation in the years 1961-2005. This minimum was fixed to only 10 years in the case of precipitation and 5 years for maximum and minimum temperatures, in order to keep most of the information of the data base. In this way, 265 pluviometric stations and 72 thermometric stations were selected.

The first exploratory analysis showed many potential inhomogeneities in the three climatic variables studied, from point errors (fig. 5), shifts in the means (fig. 6) and trends, often in the same series. Many of these apparent inhomogeneities may be caused by errors in the neighbouring stations, and therefore the more convenient strategy is to proceed iteratively, detecting and correcting only the coarser errors in the first place, and leaving the less prominent ones to the subsequent analysis.

For the first process, a big weighting parameter 'a' should be chosen (we may even set a=0 to give equal weight to all stations), to avoid an excessive influence of the nearest neighbouring station's errors on the reference series, and look for prominent point errors and shifts in the averages. Point observational errors may be easier to detect in daily data (a 10°C error in a daily lecture will only yield a 0,3°C error in the monthly average), but for the other kind of inhomogeneities it is better to avoid the higher variability of daily data and work with monthly, seasonal or even annual series.

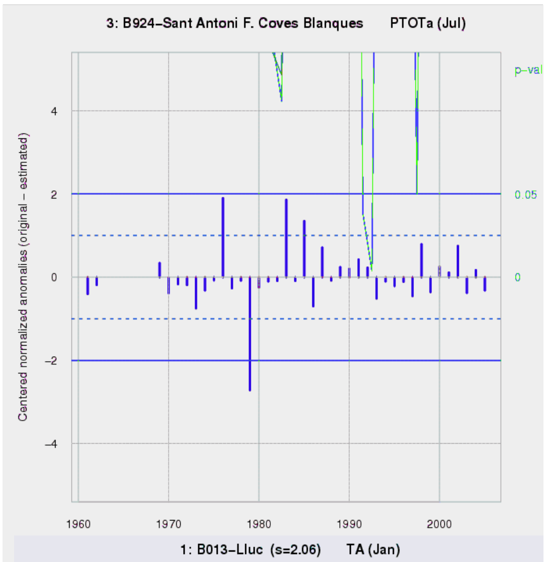


Fig. 5. Potential point errors (and a possible shift) at one station in the July precipitation series

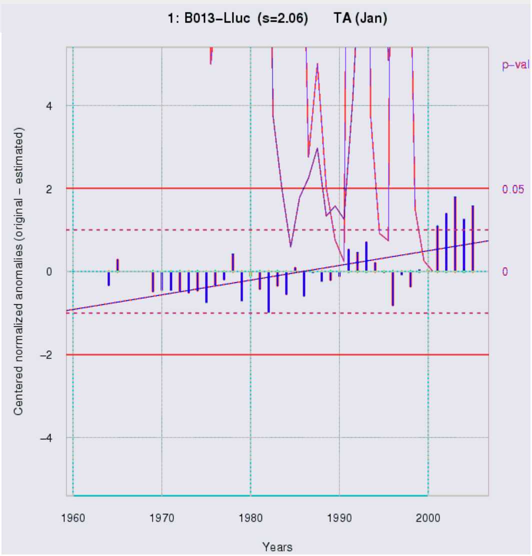


Fig. 6. Possible shifts in the mean of the monthly maximum temperatures at one station, producing also a positive trend test. Minimum p-values of the 10 (red) and 20 (green) terms moving windows act as pointers to the possible years of the shifts.

Careful attention must be paid to the peculiarities of the studied climatic variable when analysing daily data. Precipitations of convective origin will produce high spatial variability, and therefore greater differences may be acceptable between the observed and the reference series. On the other hand, maximum temperatures are easier to treat than minimums, since nocturnal inversions may produce high variability as well. In this case it is advisable to avoid overweighting near stations and to enhance the vertical coordinate in the computation of distances between stations. This kind of considerations may be also of application when studying monthly values, as can be seen in fig. 7, where the scarcity of July precipitations appears as the cause of the many potential point errors shown in fig. 5.

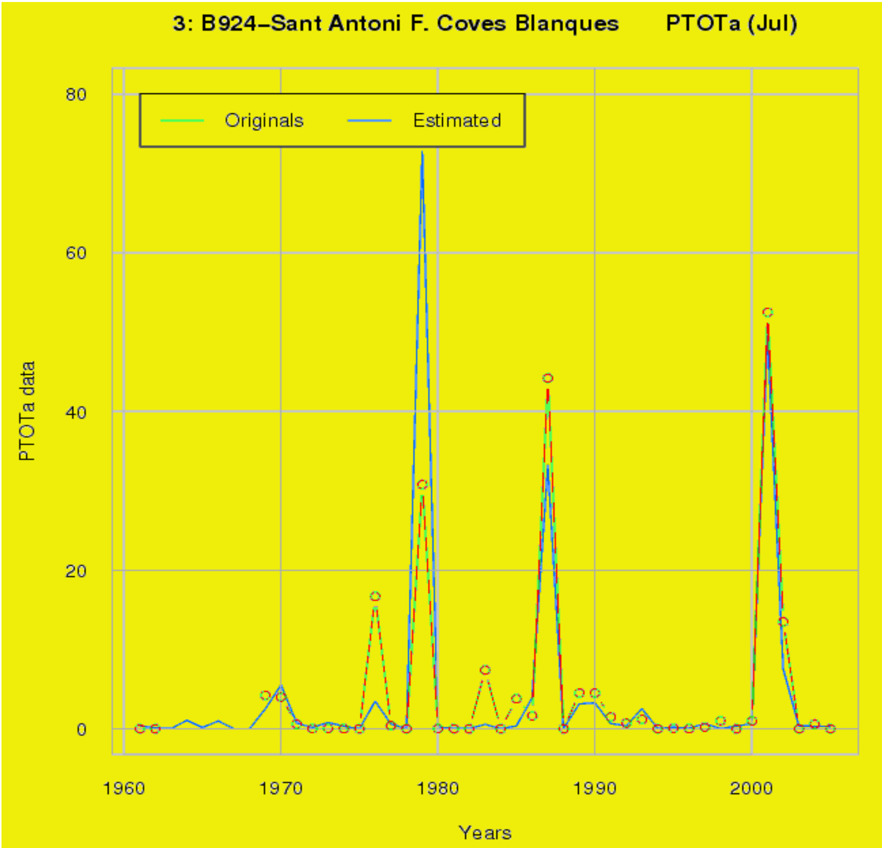


Fig. 7. Observed and reference series of July precipitation. (Same station as in Fig. 5)

Once the bigger errors have been corrected in the data base, the following analysis will allow us to better detect further point errors. When all confirmed errors are corrected or deleted, the analysis may focus on shifts of the means, that may also require repeated processes. If a shift is confirmed by the history of the station, sometimes these metadata provides us also with clues to choose which period of the series is the correct one. In this case, the other period(s) can be adjusted to remove the shift. In the contrary, it is a common practice to adjust the series to the more recent period, but unless we are quite sure of the current quality of the observations, it would be better to split the series in homogeneous intervals, and consider them as different samplings of the same location. (Spatial analysis will determine afterwards which one is more representative and which may be faulty or affected by local scale factors).

Trend analysis should be the last type of inhomogeneity to be treated, since shifts in the mean often produce positive trend tests. Afterwards, a final process may be needed if we want a data set where gaps of missing data are filled. In this last run of the program, a small value may be assigned to the weighting parameter, to enhance the information from the nearer stations. This must be given careful consideration depending on the spatial variability of the studied climatic variable, but it will have great importance when the data set is to be used for time variability studies, in order to avoid a high decreasing of the variance giving much weight to many neighboring stations.

All these processes are still been carried out for the Balearic data set, since they involve a lot of work to investigate in the data archives for all possible inhomogeneities pointed by the program. When the climatologist do not have access to the original data or have no time to accomplish these tasks, corrections can only be made on probability bases.

CONCLUSIONS

The method presented maximizes the use of the information of a climatological data set, building reference series from neighbouring stations even if they do not have any common period of observation. But as homogeneity tests in each series may be affected by errors in other nearby stations, an iterative approach of detection-correction must be undertaken, beginning by the most prominent errors.

In every of this stages, statistical tests can provide a collection of possible inhomogeneities, but they must be complemented with the visual inspection of the graphical representations of the series of anomalies. The final decision on which inhomogeneities to correct must rely on expert judgements based on knowledge of the spatial variability of the involved climatic variable and, whenever possible, on records of the history of the observatories.

REFERENCES

Aguilar E, López JM, Brunet M, Saladié O, Sigró X, López D (1999): Control de calidad y proceso de homogeneización de series térmicas catalanas. La Climatología española en los albores del siglo XXI (Raso, Martín-Vide, Eds.), Asociación Española de Climatología, A-1:15-23.

Alexandersson H (1986): A homogeneity test applied to precipitation data. J. Climatol., 6:661-675.

Almarza C, López JA, Flores C (1994): Homogeneity study of some long rainfall series in Spain: application of a new relative homogeneity test in cases of little precipitation. Climate Variations in Europe, Pub. of the Acad.of Finland

Gandin L.S. (1983): Objective analysis of meteorological fields. Gidrometeorologiches koe Izdatel'stvo, Leningrad, 242 pp.

Paulhus J.L.H., Kohler M.A. (1952): Interpolation of missing precipitation records. Month. Weath. Rev., 80:129-133.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincint L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland EJ, Hanssen-Bauer I, Alexandersson H, Jones PD, Parker D (1998): Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18:1493-1517.

Romero R., Guijarro J.A., Ramis C., Alonso S. (1998): A 30 year (1964-93) daily rainfal database for the Spanish Mediterranean regions: first exploratory study. Int. Jour. Climatol., 18:541-560.

Young K.C. (1992): A three-way model for interpolating monthly precipitation values. Monthly Weather Review, 120:2561-2569.

OBTAINING A HOMOGENIZED DATASET OF MONTHLY SPANISH MAXIMUM AND MINIMUM TEMPERATURES

M. Staudt, M.J. Esteban-Parra and Y. Castro-Díez
Departamento de Física Aplicada, Universidad de Granada, 18071 Granada, Spain, Tel. + 34-958-244023, + 34-958-294486, Fax + 34-958-243214 (Y.Castro
mstaudt@ugr.es

ABSTRACT

We present a homogenization method for the available long-term monthly series of Spanish minimum and maximum temperatures from the late 19th century on, in order to obtain a high-quality data set.

The first step is the organization of the data series (41 series) into climatically homogeneous regional groups, following a preliminary study and empirical knowledge. Then the whole set of temperature difference series is computed in each regional group, in order to perform multiple comparisons of these differences, to explore their stationarity characteristics and to detect discontinuous breaks and other inhomogeneous features. The method is based on relative homogeneity and treats absolute homogeneity only as a secondary concept, because it is generally not achievable. No reference series is used, since no reliable reference is readily available.

In the statistical analysis, the difference series are scanned with moving t, Alexandersson, and Mann-Kendall tests, under consideration of the sensitivity of these tests to the autocorrelations and in carefully chosen test intervals. An inhomogeneity is detected when several (at least three) difference series confirm a highly significant inhomogeneity. The detected inhomogeneities are adjusted by weighted averages of the regional series; these weighting factors depend on the cross-correlations and on the common data coverage.

The homogenization method is iterative and advances in steps of detection, adjustment and actualisation. Individual inhomogeneous data are discarded and gaps are filled by similar weighted means.

For posterior analysis of the temperature evolution in the Iberian Peninsula, each region is finally represented by one local series and the regional average.

Rigorous homogeneity can generally not be achieved, because the initial data quality is deficient in many cases and metadata are sparse. Nevertheless, the data homogeneity has been considerably enhanced: the total uncertainty margin in the series is of the order of 0.3ºC, under consideration of a worst-case error accumulation. On the other hand, many inhomogeneities are detected and their average amplitude is of the order of 1ºC: this number reflects the much larger error margin in the raw data. This new homogenized dataset prepares an important basis for the subsequent detection of thermal changes in Spain in the last 130 years, on a clearly higher confidence level than before.

1. INTRODUCTION

Unfortunately, a vast majority of all climate records is adversely affected by non-climatic changes in the data, due to observatory or instrument relocations, variations in the environment or in reading procedures, human errors in data processing, among others. Hence, in many cases, a series may fail to represent the real climatic evolution and a reliable detection of climate change is hard or impossible. For example, Wijngaard *et al.* (2003) analysed the daily temperature and precipitation data of the European Climate

Assessment (ECA), and found that a vast majority of the series suffer severe homogeneity problems.

There is a well-known variety of standard literature on homogenization and on the tests and methods (Goossens and Berger, 1985, Alexandersson, 1986, Karl and Williams, 1987, Young, 1993, Rhoades and Salinger, 1993, Peterson and Easterling, 1994, Easterling and Peterson, 1995, Vincent, 1998, Vincent and Gullett, 1999 and Mestre, 1999, among others). The softwares AnClim (Stepanek, 2003) and MASH (Szentimrey, 2000) were developed for the homogenization of climate data and several recent studies like Slonosky *et al*. (1999) and González-Rouco *et al*. (2001) made attempts to homogenize European data.

Nevertheless, there is still a lack of systematic homogenization treatment of long-term monthly Spanish temperature data. The present study prepares these data series, improves the data quality and tries to set a solid base for subsequent analysis of thermal changes on a regional scale since the late 19th century.

2. DATA

The monthly Spanish temperature records by the National Meteorology Institute (INM) provide data from 41 observatories with a minimum coverage of 30 years, as illustrated in Figure 1 (about half of them include 19th century data, beginning between 1869 and 1880). The data quality is problematic or sometimes even poor, because of frequent site changes and data gaps, and metadata are scarce. Figure 1 shows the geographic distribution of the observatories.

Definition of the regional groups of data series

The Spanish monthly temperature series contain a high degree of common variability: the cross-correlation coefficients between the anomalies usually exceed 0.5. Nonetheless, in spite of this dominant common variability pattern (the "peninsular mode"), the temperature anomalies show certain regional distinctions.

Based on a previous regional analysis, we organized the data series into the following climatically different groups: the central plains (14 series), the Mediterranean (6 s.) and the Cantabrian (5 s.) coastal areas, Galicia (6 s., the northwest), western Andalusia (4 s., basically the Guadalquivir valley), Extremadura (2 s.) and the Ebro valley (4 s.). The aim was to preserve possible regional distinctions through a regional homogenization.
In each of these climatic regions, all the series were homogenized, then the regional mean series (simple mean of the anomalies) were computed a-posteriori, to represent the region, together with one chosen individual series.

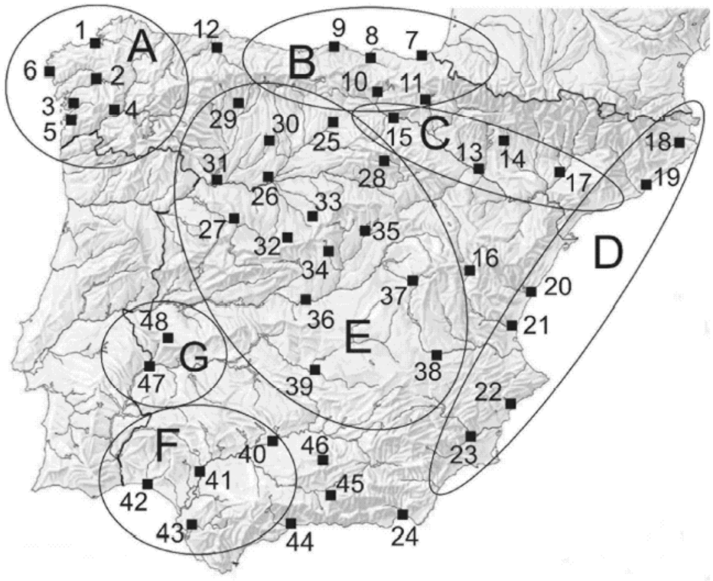


Fig. 1. The spatial coverage of the Spanish maximum and minimum temperature series, between 1860 and 1980 (about half of the series include 19th century data). The series are: 1. La Coruña, 2. Santiago, 3. Pontevedra, 4. Orense, 5. Vigo, 6. Finisterre, 7. San Sebastián, 8. Bilbao, 9. Santander, 10. Vitoria, 11. Pamplona, 12. Oviedo, 13. Zaragoza, 14. Huesca, 15. Logroño, 16. Teruel, 17. Lérida, 18. Gerona, 19. Barcelona, 20. Castellón, 21. Valencia, 22. Alicante, 23. Murcia, 24. Almería, 25. Burgos, 26. Valladolid, 27. Salamanca, 28. Soria, 29. León, 30. Palencia, 31. Zamora, 32. Ávila, 33. Segovia, 34. Madrid, 35. Guadalajara, 36. Toledo, 37. Cuenca, 38. Albacete, 39. Ciudad Real, 40. Córdoba, 41. Seville, 42. Huelva, 43. Jerez, 44. Málaga, 45. Granada, 46. Jaén, 47. Badajoz, 48. Cáceres. The regional groups are: A Galicia, B Cantabria, C Ebro valley, D Mediterranean, E central plains, F western Andalusia and G Extremadura.

3. OUTLINE OF THE HOMOGENIZATION METHOD

A. Some statistical properties of monthly temperature data

The statistical distribution of temperature data is normal as a good approximation and we can apply parametric statistics designed for Gaussian-distributed variables, as the t-test or the SNHT. The autocorrelations in these series are rather slight (coefficients between 0.1 and 0.3), but several statistical tests require corrections (the reduced sample size for the t-test and prewhitening of the series for the Mann-Kendall test), in order to achieve realistic confidence levels.

B. The homogenization concept

If a series is homogeneous in an absolute way, there is no variability, except for the real climatic evolution. However, this condition is almost never fulfilled (Easterling et al., 1996, pointed out that "… *the real homogeneity of climatic data is irretrievably lost*") and we generally cannot decide, through an analysis of just one series, at a good confidence level, whether or not a certain change is inhomogeneous.

Therefore, we did not follow an absolute homogeneity method, but a relative homogeneity concept, based on difference series: the anomalies of highly correlated time series are essentially synchronous and its differences should be approximately random. A local inhomogeneity can be detected in these differences, but a real extreme anomaly tends to vanish. This detection method fails if several series suffer a simultaneous data problem (e.g. a common sudden jump). Comparing as many difference series as possible minimizes this risk.

Our homogenization method is based on multiple comparisons between the series within each predefined region. We do not work with regional reference series, because the rather low number of series with frequent inhomogeneities does not permit the creation of a reliable *a priori* reference. The whole set of difference series (differences of anomalies) is statistically tested for significant changes. Once identified, an inhomogeneity is adjusted by a weighted mean of the highest-correlated series. The weighting factors depend on the synchronicity (cross-correlation) and the number of common data.

We have not avoided to reject and to delete clear outliers (see C.3), intervals (shorter than a decade, with more than one third of missing data and disconnected intervals, without any appropriate information to connect them to the rest of the series) or even whole series (with more than five clear inhomogeneities), when the homogeneity problems were too strong for an adjustment at an acceptable confidence level.

C. The homogenization scheme

1. We converted the raw-data series into anomalies (relative to a recent reference interval) and computed the whole set of anomaly difference series within each region.
2. We marked the suspicious inhomogeneities (abrupt changes or breaks, outliers), with particular attention to the metadata.
3. We discarded the largest and most obvious outliers that exceeded a certain level, based on the difference series (four standard deviations of a running 30 year-interval). The severe criterion of this preliminary step removed only the very large inhomogeneous outliers.
4. After recalculating of the set of difference series, we searched for abrupt inhomogeneities (breaks) and then defined individual appropriate “base intervals” for their statistical detection. Where it was possible, the length of a base interval was chosen as 20-30 years, around the possible break point[1]. We had to avoid strictly a temporal overlapping between inhomogeneities and their intervals, but tried to achieve a reasonable sample size (at least of the order of 100).
5. The moving t and SNHT (Alexandersson) tests[2] were applied to the whole set of difference series in the base intervals of point 4, scanning the intervals, to determine the probability of a break point, as a function of its time. We examined first the intervals around the incidents reported in metadata. An inhomogeneous break was detected when the significance level exceeded 99% in the t-test and was at least 50% above the 95%-level in the SNHT, recurrent in three difference series[3]. We checked the local anomalies and, in doubtful cases, ran also the sequential Mann-Kendall-test (with “prewhitened“ series).

[1] We considered the so-called “station drift” since the differences between correlated temperature series show frequent trends of changing signs (even in absence of inhomogeneities, see Rhoades and Neill, 1995). Therefore, the stronger this drift is, the shorter the base intervals of the candidate series must be, , because earlier or later data are less valid for the adjustment at a certain time.

[2] We corrected the significance levels of the t-test for autocorrelations by the” reduced sample size”.

[3] The t-test generally confirmed the results of the SNHT; the latter has sharper peaks, due to its quadratic algorithm, see Figure 2.

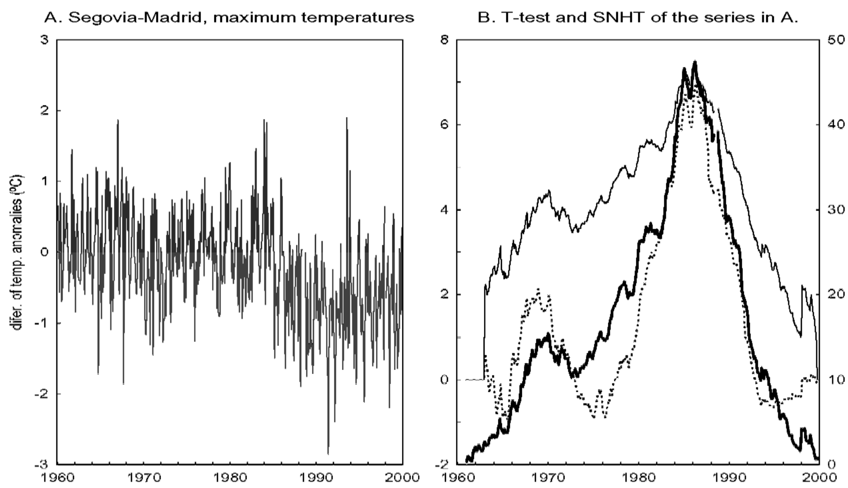


Fig. 2. A: a difference series of maximum temperatures; B: the coefficients of the t-test with a 20-yr running window (discontinuous line) and in the whole 40-year interval (both left axis) and of the SNHT in the 40-year interval (thick line, right axis).

6. After the detection, we adjusted an inhomogeneous break with the highest-correlated simultaneous regional data (up to five series, see Figure 3). The candidate series' *after-before* difference (offset) was replaced by a weighted average of the synchronous differences of the correction series. The weighting factors were given by the squared cross-correlations and the common data fraction of each series, relative to the candidate. In few particular and highly significant cases, we detected continuous inhomogeneities and adjusted them in a similar way[4]. We always adjusted the data before a break.

To assure the non-interference between the adjustments, we performed the steps 4-6 iteratively: after adjusting all the disjointed inhomogeneities, we recalculated (updated) all series and went on with the next iteration. In many cases, slighter inhomogeneities could be detected only after adjusting a large inhomogeneity. The iteration stopped when no more significant inhomogeneities were detected.

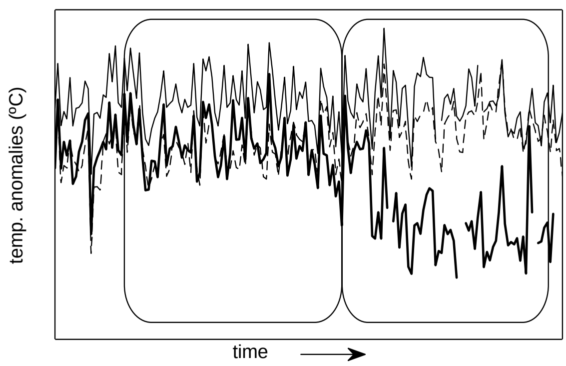


Fig. 3. An illustration of the adjustment method. The thick line represents the anomalies of the candidate series with a detected break between the two marked adjustment intervals.

[4] In some cases, when there is a sufficiently long overlapping interval between two candidate subseries, we verified their synchronicity and absence of inhomogenities and adjusted by the mean difference.

7. We searched for individual inhomogeneous data, by detecting extreme values of the difference series and verified these data in each local series (in analogy to point 5, we considered an outlier as inhomogeneous when its amplitude exceeded the 99.5% confidence level in at least three difference series, relative to a symmetrical running 30-year-interval). The detected inhomogeneous data were removed.
8. The missing data (or gaps due to removed inhomogeneous data) were filled by weighted means of the best-correlated synchronous data, based on standardized anomalies in adjacent years, assuming synchronicity between these series. The filling algorithm used up to five regional series and weighted the contribution of each time series, as in point 5, with the squared cross-correlations and a common data factor[5]. When the information of the surrounding series were insufficient, the gaps were filled by an ARIMA- interpolation (only intrinsic information of the candidate series).
9. Finally, the dataset was prepared with two time series for each climatic region, expressed as anomalies, relative to the reference period 1961-1990: one local series and the regional average series (all the local series are also available, for further purposes).

Seasonally or monthly varying adjustments?

The adjustments were based on the monthly data, but generally did not vary with the month or the season (only a few seasonally distinct adjustments were applied, when the seasonal discrepancies were particularly large). In literature, we find different types of adjustments (see, for example, Peterson and Vose, 1998). Inhomogeneities in climate data often depend on the month or season, because of the seasonally diverging impacts of instrumental or environmental changes and under this viewpoint, a monthly or seasonally varying adjustment is theoretically better. However, it modifies the variability, the autocorrelation structure, and the annual cycles of the data, whereas the adjustments applied here consist of a simple additive term. Furthermore, a monthly varying adjustment must work with 12 times fewer data (for a given interval length) and the confidence margins are substantially wider. Hence such an adjustment becomes more attractive when the number of series is larger and when the initial data quality is higher than in the present study.

4. RESULTS

A. The homogenized dataset, adjustments and rejected data

Among the 43 monthly maximum and minimum temperature series, we found widespread homogeneity problems: adjustments were necessary in almost all series, although the criteria for the detection of inhomogeneities were rather severe (high significance and redundancy levels). In some cases, long intervals or whole series were rejected, because of a lack of homogeneity. We adjusted a total number of 59 (85) inhomogeneities in maximum (minimum) temperatures (see Table 1), with a mean amplitude of 1.00ºC (1.05ºC), besides the rejected intervals and individual data. As an average, we applied one adjustment every 44.5 years (66 years).

[5] The amplitude of the correction is reduced (filled value closer to the mean of the chosen interval), when the information of the nearby series drops below a certain threshold.

B. The error margins of raw and homogenized data

The instrumental error in temperature measurement is of the order of 0.1ºC (Linacre, 1992) and increased to around 0.2ºC in differences between two series. A series of roughly one century required an average of two adjustments (see point A); therefore the mean margin of the total error (instrumental plus homogenization, under a worst-case error accumulation) increased to about 0.4ºC.

On the other hand, the mean amplitude of the adjustments (around 1ºC) defined the uncertainty in the raw-data series (mean error due to the inhomogeneities), besides the instrumental error. A long series had between one and five inhomogeneous breaks, with an average of around two. Depending on the degree of cancellation of these errors, the total uncertainty in the raw series frequently exceeds 1ºC, and sometimes it could even be higher than 2ºC.

C. A comparison of some results, based on raw and homogenized data

We compared the temperature changes with raw and homogenized data, by applying a t-test to the means of different intervals and regions. In some cases, an inconsistent result has been found between the raw local and mean series (only one of both series showed a highly significant change). After homogenization, this inconsistency between the highly correlated regional representative series disappeared, indicating a higher degree of redundancy between the series and therefore a better data quality.

Table 1. Total data and number of adjustments and rejected data (individual data or intervals, y = years) in the Spanish maximum and minimum temperature series.

Series	*No. data*	*No. adj.*	*rej. data*	*No. data*	*No. adj.*	*rej. data*
	Maximum temperatures			**Minimum temperatures**		
La Coruña	1460	1		1460	1	
Santiago	1066	1	4	1081	1	1
Pontevedra	1064	1	15	866	2	236
Orense	951	2	76	452	1	88
Vigo	1002	2	72	876	2	146
Finisterre	524	1	1	510	1	5
San Sebastián	1485	1		1485	2	
Bilbao	768	2	82	963	1	1
Santander	1076	2		1075	1	
Vitoria	979	1	1	980	3	2
Pamplona	1370	2	29	1126	1	≈ 34 y
Zaragoza	1592	2		1592	3	4
Huesca	1542	2	9	1525	2	4
Logroño	958	2	1	969	3	
Lérida	607	2	160	699	2	50
Valencia	1592	3	4	1592	3	≈ 25 y
Gerona	1072	2		1066	5	6
Barcelona	1496	2	1	1095	3	≈ 26 y
Castellón	1049		1	1038	2	16
Alicante	-	-		1157	2	≈ 31 y
Murcia	1592	2	1	1430	4	77

Madrid	1592	3	2	1592	2	5
Ávila	929	2		1006	1	20
Burgos	1556	2	3	1558	2	5
León	432		≈ 21 y	923	2	
Palencia	873	1		877		1
Salamanca	1558	1		1459	6	8
Segovia	1401	1		1399	4	3
Soria	1535	2		1365	3	≈ 6 y
Zamora	790			1038	1	3
Guadalaj.	893	2	1	313		≈ 48 y
Toledo	1101	1		1085	1	17
Albacete	1366	1		1418	4	8
Ciud. Real	1092		2	--	--	--
Cuenca	1074		1	692		174
Seville	1220	3	2	1220	4	2
Córdoba	1048		2	1050	2	3
Huelva	1046	2	2	1005	2	171
Jerez	754	1	≈ 24 y	919	1	7
Málaga	1114	2		1302	2	
Badajoz	1112	2	1	1112	3	
Cáceres	1122		7	1000		2

5. CONCLUSIONS AND DISCUSSION

Thermal changes are expected to be generally of the order of 1ºC or smaller; hence, the large error margin of the raw Spanish long-term data frequently does not allow a reliable detection of changes of this order. This situation is largely improved by the homogenization method: it applies severe conditions for inhomogeneity and may fail to detect small inhomogeneities, but the average total error of the series drops below half a centigrade and clearly enhances the data quality for subsequent analysis (and allows to a certain degree an analysis of the regional differences, besides the first-order trends in Spain). Furthermore, we confirmed an improvement of consistency between highly correlated series, due to homogenization.

This homogenization method was developed for a rather small number of series of limited quality, where the level of data redundancy is sometimes low and the intervals free of inhomogeneities are frequently short (shorter than 20 years and often shorter than 10 years). Under these circumstances, we applied multiple comparisons, instead of a method with reference series, and generally we did not perform monthly or seasonally varying adjustments. We consider this method a kind of “first order homogenization” (or a method for low data coverage): its results are significantly better than those of the “zeroth order” (the raw data), but the method may be refined, when the data coverage and/or the initial data quality is substantially higher.

REFERENCES

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661-675.

Easterling DR, Peterson TC. 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369-377.

Easterling DR, Peterson TC, Karl TR. 1996. On the development and use of homogenized climate datasets. *Journal of Climate* ***9**: 1429-1434*.

González Rouco JF, Jiménez JL, Quesada V, Valero F. 2001. Quality control and Homogeneity of Precipitation Data in the Southwest of Europe. *Journal of Climate* **14**: 964-978.

Goossens Ch, Berger A. 1986. Annual and seasonal climatic variations over the northern hemisphere and Europe during the last century. *Annales Geophysicae* **4B4**, 385-400.

Karl TR, Williams CN. 1987. An Approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology* **26**: 1744-1763.

Linacre E. 1992. Climate Data and Ressources. A Reference and Guide. Routledge, London and New York, p. 35.

Mestre O. 1999. Step-by-step procedures for choosing a model with change-points. Proceedings of the second seminar for homogenization of surface climatological data. (Budapest, Nov. 1998), World Climate Programme, Data and Monitoring, World Meteorological Organization.

Peterson TC, Easterling DR. 1994. Creation of homogeneous composite climatological reference series. *International Journal of Climatology* **14**: 671-679.

Peterson TC, Vose R, Schmoyer R, Razuvaëv V. 1998. Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *International Journal of Climatology* **18**: 1169-1179.

Rhoades DA, Salinger MJ. 1993. Adjustment of temperature and rainfall series for site changes. *International Journal of Climatology* **13**: 899-913.

Rhoades DA, Neill WA. 1995. The effect of station drift on the estimation of shifts in meteorological records due to site changes. *International Journal of Climatology* ***15**: 207-220.*

Slonosky V, Jones PD, Davies TD. 1999. Homogenization techniques for European monthly mean surface pressure series, *Journal of Climate* **12**: 2658-2672.

Stepanek P. 2003. AnClim: a software, specially developped for the homogenisation of climatic data: webpage http://www.sci.muni.cz/~pest; contact: pest@sci.muni.cz or petrstep@email.cz.

Szentimrey T. 2000. Multiple Analysis of Series for Homogenization (MASH). Seasonal Application of MASH (SAM), Automatic Using of Meta Data. Third seminar for homogenization and quality control in climatological databases. Budapest, September 25–29. Hungarian Meteorological Service.

Vincent LA. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11**: 1094-1104.

Vincent LA, Gullett DW. 1999. Canadian historical and homogeneous temperature datasets for climate change analyses. *International Journal of Climatology* **19**: 1375-1388.

Wijngaard JB, Klein Tank AMG, Können GP. 2003. Homogeneity of the 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23**: 679-692.

Young KC. 1993. Detecting and Removing Inhomogeneities from Long-Term Monthly Sea Level Pressure Time Series. *Journal of Climate* **6**: 1205-1220.

QUALITY CONTROL OF MONTHLY PRECIPITATION DATA FROM MEDITERRANEAN AREAS OF SPAIN

José Carlos González-Hidalgo (1), Martín de Luis Arrillaga (1), Petr Stepanek (2), Siham Lanjeri (1).
(1) Dep. Geography, University of Zaragoza, Spain.
(2) Hydrometeorological Service, Brno División, Chezck Republic.
Dep. Geography, University Zaragoza, 50009, Zaragoza, Spain
jcgh@posta.unizar.es

ABSTRACT

We developed quality control (anomalous data detection and homogeneity analysis) and reconstruction process of all monthly precipitation data for eastern Mediterranean area of Spain, stored at Nacional Meteorological Agency (Instituto Nacional de Meteorología). The original amount of observatories is 3891 and more than 10^6 monthly data represent the total precipitation dataset. Due to difficulties in creating suitable reference series, the ones covering less than 10 years were discarded. Data duplications in different series of the same location has been also checked and removed. In consequence of these two previous steps, 5.5% of original data was deleted and altogether 2669 series were analyzed (containing 964.173 monthly data). AnClim and ProClimDB Software were used in the analysis.

For each monthly series (2669*12), we calculated an independent reference series using all close-by neighbours (less than 50 km apart) with a minimum overlap period of 10 years; positively correlated and with a mean monthly correlation > 0.5. With these restrictions, reference series were calculated by weighted mean $(1/\text{distance})^2$ after mean standardization.

Suspicious data were detected by comparing the ratio series (candidate / reference, and viceversa, because of minimum precipitation value is set to 0). This procedure was combined with inter quartile ranges. To avoid influence of suspicious data, we followed an iterative process (10 steps). Then, after removing the first set of suspicious data, a new reference series were calculated and all data were checked again. At the end, altogether 7182 data (0.75 % of total data) has been removed . After that, we checked for homogeneity in all series by applying SNHT. To test the influence of anomalous data in homogenization process, we run SNHT test both using original data-set and data-set obtained after anomalous data elimination. Thus, in original database a total of 1966 non homogeneous series (75%) (containing 2984 inhomogeneities) were detected. However after applying SNHT on depurated data base, only 1125 series were detected as non homogeneous (43%), including 1795 inhomogeneities.

The results indicate that detection and correction of suspicious data in huge databases appears to be necessary in order to avoid statistical inhomogeneities. In conclusion we present some initial results consisting of monthly precipitation trends during the winter months.

1. INTRODUCTION

In the last report of the IPCC, sub-regional analysis of the global change, precipitation analysis and need of articulating global, regional and local scale of climate were considered as reliable objectives (Houghton et al., 2001; Parry, 2001). These tasks suppose that results of General Circulation Models (GCMs) and empirical data are linked. (Allen and Ingram, 2002), and, usually present two difficulties. First, the precision and reliability of GCMs decrease with the change of scales (Matyasovszky et al., 1999; Mcnamara, 1999). By using 3º x 3º longitude/latitude resolution, validations are well accepted in extensive and homogenous areas. Nevertheless, until now, GCMs are not able to catch the spatial shades of regional and local scales (Zorita and González Rouco, 1999; Prudhoumme et al., 2002). The phenomenon is mainly observed in precipitation analyses (Allen and Ingram, 2002), especially in areas where the pluvial regime is characterized by shortage, seasonality and variability, as in areas of Mediterranean climate (Groisman and Legates, 1994; Balairón, 2000; Wilby and Wigley, 2000). In this context, predictions of GCMs represent a simplified vision of the reality not well fitted to the space details (Palutikof et al., 1996; Sulzman et al., 1995), and of increasing uncertainty in temporary scales (Mearns et al, 1995; Barrow et al., 1996), which needs to be confirmed by means of study of historical data.

On the other hand, comparison of weather stations' data with GCM results requires the availability of detailed spatial and temporal information. The situation particularly affects precipitation, being one of the most variable climate elements. In addition, this information is fundamental not only in evaluating GCMs, but also in comparing observed climate change to natural variability (Hulme et al., 1995; New et al., 2002).
Concerning the first aspect, and to facilitate the comparisons, it has been suggested that the data analysis must be made on a minimum (about 30 years long) period of observations, which corresponds tothe normal periods of the World Meteorological Organization (Hulme et al., 1995; Moron et al., 1995). Although, other periods has been suggested as well, like 1951-1980, which allows the analysis of the NAO index (Hurrel, 1995; Rodrigo et al., 2000). The second question concerns the precipitation changes, which can only be detected when a dense network of observatories is used (Groisman and Legates, 1994; Hulme, 1995; Vinnikov et al., 1990; Cosgrove and Garstang, 1995).

The previous comment implies two conclusions: (i) We must be extremely careful in areas of climatic transition when analysis of climate change is required on sub-regional scales. (ii) In this case, we must avoid spatial and temporal generalizations, and for these reasons, the study of the observatories data becomes essential.

The availability of dense and prolonged databases is very diverse, depending on the areas. In Europe the University of East Anglia made the biggest effort (among other institutions) (Hulme et al., 1995; Tank et al., 2002), while at sub-regional scale, disparity is great. In fact, in the Iberian Peninsula and Spain, several studies have used database gathered by observatories of greater quality and duration, and provided general view on the evolution of precipitation (Esteban-Parra et al., 1998; Rodriguez-Puebla et al., 1998; González-Rouco et al., 2001, among others), or its relation to tele-connection patterns (Martín Vide, 2001). At the same time they were not able to detect sub-regional scale characteristics, as suggested by the IPCC, because the spatial density of observations is very low (around 1 observatory in 5000 km^2). Therefore, general models calibrated on these databases can also offer interesting results, although their resolutions are not good enough to investigate the effects of climate change.

Consequently, database construction and analysis of historical data appear to be a very promising way to make possible global change studies at different scales. In addition, it seems to be particularly adaptable in areas where climatic transition takes place, and probably, effects of the climate change will be well perceived. The Mediterranean climate areas belong to this universe, and the Iberian Peninsula is a good example.

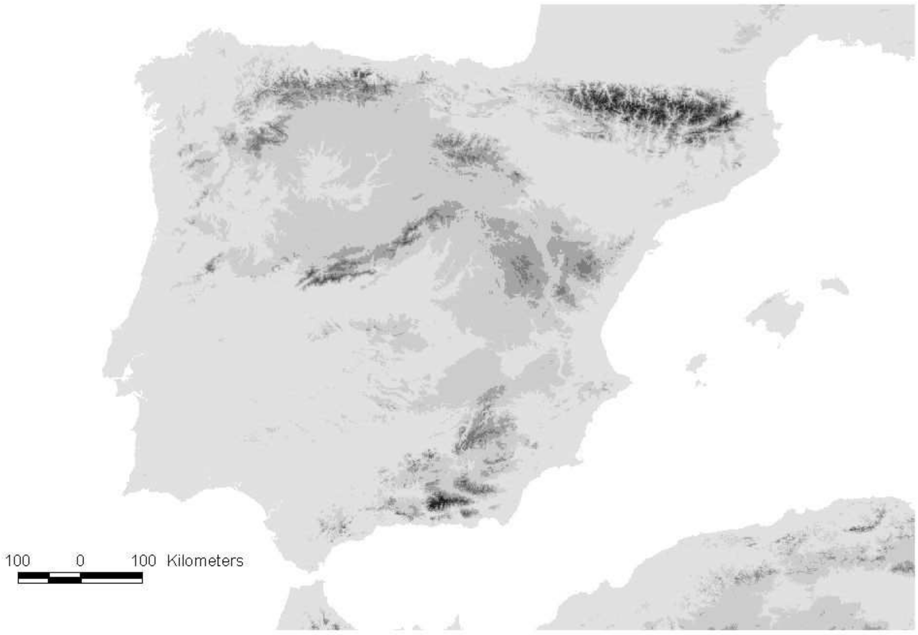


Fig. 1. Spatial distribution pattern of Spanish principal mountain chains

The Iberian Peninsula cover a total area of about 500000 km2 and it is located in the western part of the European continent, within the latitudinal strip of 36º-44º north. Its location is peculiar mainly due to its relief. In the Iberian Peninsula the main mountains are distributed in parallel strips W-E from north to south (Cantabrian, Pyrenees, Central System, Andalusian mountainous areas), which are blocked to east by the Iberian System in direction of NW-SE-S-SW (Figure 1). Therefore, latitudinal gradation of N-S takes place in the pluvial values, overlapping with another gradation of W-E. The Mediterranean areas, approximately 1/3 of the total area of Spain, are safe from the influence of the Atlantic, and has their own personality in which rainiest places are detected (Mountain range of Grazalema, province of Cadiz), as well as a unique European desert (Cabo de Gata, province of Almeria).

In this paper, we present the development of methodological approach proposed in the previous seminar celebrated in Budapest (González-Hidalgo et al., 2004), on techniques and control quality of climatic data. The study leads by the analysis of dense data base offered by the National Institute of Meteorology (INM) of Spain. Our purpose is to test different methods of analyses, and here we present some provisional results for precipitations trends.

2. DATA BASE

The database have been collected from the total records stored at the National Institute of Meteorology of Spain, and present all the available precipitation data recorded in the study area (3891 observatories including more than 10^6 monthly data). The series are highly variable in characteristics (gaps, length, etc) and several different series exist at the same location (from different observatories), or at sites located close to each other (< 5 km). Furthermore, spatial density of observatories is high, except for altitudes higher than 2000 m o.s.l.

This data base have been analyzed partially at global scale using the highest quality observations (González Rouco et al., 2001), or at sub-regional scale (e.j. Sumner et al., 1993; Pérez-Cueva, 1994; Lana et al., 2004). Consequently, global study gives an opportunity to obtain hundreds of series, at least from the second half of the 20th century. We are consistent in pointing out the absence of other studies leading with quality control at the same scale as raised in this work.

3. METHODS

The quality control applied consists of four steps: identification of repeated data within the same series, identification of repeated series, detection of anomalous data and control of homogeneity.

3.1. Identification of repeated data within the same series

Different situations were detected in analyzed dataderiving from the same series. Consecutive repetition of identical monthly data, repetition of monthly data in consecutive years, and chains of zeros in four successive years. Table 1 shows some selected examples of the above situations. These data have been eliminated.

Table 1. Selected examples of repeated data within the same series (in dec. mm)

Code	Yr	J	F	M	A	M	JN	JL	A	S	O	N	D
0220	1968	1	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0220	1969	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0220	1970	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0220	1971	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
0347c	1982	**1096**	**1631**	**930**	**452**	**490**	**468**	**720**	**2015**	**371**	**753**	**1353**	**8**
0347c	1983	**0**	**374**	**100**	**452**	**495**	**468**	**720**	**2015**	**371**	**753**	**1353**	**8**
0347E	1982	**1096**	**1631**	**930**	**452**	**495**	**468**	**720**	**2015**	**371**	**753**	**1353**	**8**
0347E	1983	**0**	**374**	**100**	**452**	**495**	**468**	**720**	**2015**	**371**	**753**	**1353**	**8**
8163	1962	230	326	**1015**	**1015**	**1015**	**1015**	**1015**	**1015**	**1015**	**1015**	569	331
8200b	1975	**0**	**230**	**570**	1020	280	**340**	**530**	**400**	**280**	**700**	**0**	**0**
8200b	1974	**0**	**230**	**570**	1250	210	**340**	**530**	**400**	**280**	**700**	**0**	**0**

3.2. Identification of repeated series.

For each series, a set of neighbours have been identified by distance calculations. Then, each one has been compared to the nearest to verify its own identity and to avoid duplicity.

The procedure involved detection of numerous situations in which, up to four series apparently independent, data are fully or partially repeated. In these cases, we preserved the longest series with the most recent data, and the missings of selected series were completed by calculating average of the neighbour's data.

Finally, all series of less than 10 years have been joined to their neighbours, by using the procedure described before. Table 2 shows an example of repeated series among 8 years.

Table 2. Selected examples of repeated series (in dec. mm)

Code		J	F	M	A	MY	JN	JL	A	S	O	N	D
0003	1951	810	78	1077	1138	1038	432	67	199	2405	2623	500	1295
0003	1952	255	120	559	630	960	503	50	85	430	586	70	197
0003	1953	0	0	1069	208	364	1272	240	0	580	1612	234	1076
0003	1954	130	650	965	311	796	740	250	83	329	52	260	750
0003	1955	1383	617	157	0	180	1593	224	490	874	1069	205	867
0003	1956	740	222	1890	463	1264	495	13	385	1060	425	605	50
0003	1957	16	50	0	710	2281	1696	0	131	260	1821	365	1695
0003	1958	665	0	313	250	0	110	0	120				
Code		J	F	M	A	MY	JN	JL	A	S	O	N	D
0002I	1951	810	78	1077	1138	1038	396	67	199	2355	2621	500	1295
0002I	1952	255	120	559	630	960	503	50	85	430	586	70	197
0002I	1953			1063	208	364	1270	240		580	1612	234	1736
0002I	1954	130	650	965	311	798	740	250	83	329	52	260	750
0002I	1955	1383	617	157		180	1593	224	473	891	1002	272	868
0002I	1956	740	222	1875	468	1264	495	13	385	1060	425	605	50
0002I	1957	16	50		710	2471	1796		131	260	1825	365	1695
0002I	1958	765		313	250		110		120				

By applying such procedures of quality control, data base have been reduced to 2669 series and 964.173 monthly data, and metadata file have been created with repeated series and eventually possible combination of series. Depurated data represent 5% of the originally existed data base.

In the following steps, detection of anomalous data and control of homogeneity were made by means of reference series calculation and iteration procedure. This process is really relevant and it is appropriate to make a detailed analysis of how the reference series were obtained.

3.3. Reference series calculations

Reference series is a combination of the neighbours better correlated with the series which has to be analyzed. Therefore, it becomes a sample to which our candidate may be compared. The selection of the neighbouring observatories is based on two criteria: the distance and the correlation coefficient.

The neighborhood seems to be suitable criteria allowing for the unavailability of studies which specify maximum or minimum distances in the selection. The selected distance may depend on the topography and climatic behavior of the regions. On the other hand, absence of topographic barriers or elements that weaken the pluvial rates between the neighbouring observatories and the candidate is recommended (Vincent and Gullet, 1999).

To avoid slants of national character (type of instrument or own technique of meteorological services), it has been suggested to include always some foreign observatories (Peterson and Easterling, 1994).

Regarding correlation coefficient, positive thresholds interval from r = 0.7 to r = 0.8 have been recommended (e.j. Vincent and Gullet, 1999). Nevertheless, in areas with Mediterranean climate, where the precipitation is extremely variable throughout years and is concentrated only on a few days, even adjacent observatories do not show highly correlated values (Rodriguez et al., 1999). In fact, it results from the local character of the phenomenon, induced by the high frequency of convective processes. On the other hand, due to the dichotomizing character of precipitation (if it rains or not), and to the absence of negative values, it has been proposed to select correlations by using the difference series (Peterson et al., 1990a), or transformed logarithm (Rhoades and Salinger, 1993). This process, in addition, would avoid the outliers effects as indicated by Lanzante (1996) and González-Rouco et al. (2001). Consequently, combination of both criteria could be much more effective in this kind of study.

The second question needs to be determined for calculations of reference series is the number of selected neighbours. By Peterson and Easterling (1994) it was suggested to be around five but never lower than two. However, other studies indicated that even a single observatory might be suitable as reference one, whenever its quality is well verified (Keiser and Griffiths, 1997).

Finally, procedures for the calculation of any reference series were object of great attention in many researches (Jones and Conway, 1997; Jones and Hulme, 1996). In all cases, the weighed values of neighbouring data are often considered.

According to the above considerations, reference series were made after the process of neighbours' selection within the range of 50 km, with a minimum overlapping of 10 years, and with all positive monthly correlations with an average monthly correlations higher than 0.5.

As different neighbouring series could show different overlap periods with the candidate one, in the presence of missing data and different averages of the neighbours, some standardization process is needed in order to avoid introducing uncontrolled slants during the creation of reference series. Thus, in each case, average standardizations has been applied to all neighbour series by using common overlap period with the candidate one.

Thereafter, the calculation of each reference series was carried out by means of weighed average of $(1/d)^2$, where d is distance in km. The selection of the distance as weighted factor is required in case of precipitation, owing to the spatial character of this element. The distance of chosen selection (50 km) is well adapted to our purposes, as different tests, performed on neighbours within 10 km and 75 km, have proved.

3.4. DETECTION OF ANOMALOUS DATA

Once reference series have been calculated, the detection of suspicious data was made by means of an iterative process of detection-elimination where, the reference series was calculated again in each step. At the detection of suspicious data we considered that the precipitation appears as a series of data limited by its base (minimum value equal to zero). Thus, to avoid this effect, we analyzed series of direct ratios (C/R) and inverse ones (R/C). In the first case, series of ratios was used to identify possible anomalous data by excess in the candidate series front of its reference while in the second one, we detected possible anomalous data by default.

The detection of suspicious data, in each case, was made by combining two criteria, in which thresholds of both ratios and the inter-quartile distances have been combined, by using the expression of Q75 + (Q75 - Q25) * 3, where Q is the corresponding quartile and 3 is a coefficient.

To avoid dificulties of ratio calculations for zero values, all data base and reference series were increased by 1. Furthermore, in order to avoid the slant that lowest values could induce in the smaller ones (e.g. ratio of 10/1 = 10, ratio of 1000/100 = 10), all data were also increased by a constant of 29 mm. We used this final threshold value (29+1), since months with 30 mm of monthly precipitation can be considered as dry month in the climatic tradition (see Köppen classification). Finally, the threshold values of ratios and inter-quartile distances were applied strictly in case of zero values (Table 3).

Table 3. Threshold value for ratio and inter-quartile distance.

Normal data Threshold		Cero data Threshold	
Ratio	Quartile	Ratio	Quartile
>4.5	All	>4.0	All
> 4.0	> 2.5	> 3.5	> 2.5
> 3.5	> 5.0	> 3.0	> 5.0
> 3.0	> 7.5	> 2.5	> 7.5
> 2.5	> 10.0	> 2.0	> 10.0

The procedure was repeated successively, removing in each iteration the suspicious data from the original database and proceeding again to calculate new reference series. The final step consisted of removing all suspicious data from the initial data base, hence calculating the most depurated reference series.

These last reference series, obtained after 10 iterations, were finally contrasted with the original series in order to perform the final detection of all the doubtful data. Thus, finally a total of 7182 monthly data were eliminated, which is less than 1% of the original data base. In Table 4 and 5 we show some cases of detected and eliminated data that appeared to be suspicious. In Table 4, examples can be seen of data of candidate series with values superior to those of reference ones, in Table 5 inversely. The Candidate series C has been compared to the Reference one R, in the first case, by means of a direct ratio C/R (Table 4), adding to each value the constant of 300. In table 5, we present the contrasted values of series of ratio R/C. IDQ means the inter-quartile distance value for both ratios series, respectively.

Table 4. Selected anomalous data for C > R

Code	Year	Nº neighb.	Month	C (1/10 mm)	R (1/10 mm)	C+300	R+300	Ratio	IQD
0313	1918	5-9 st.	12	2605	19	2905	319	9.1	31.8
6119E	1988	48-66 st.	12	2808	30	3108	330	9.4	47.3
7247	1938	1-10 st.	8	2650	50	2950	350	8.4	73.8
8274U	1974	45-51 st.	12	3730	0	4030	300	13.4	143.7

Table 5. Selected anomalous data for C > R

Code	Year	Nº neighb.	Month	C (1/10 mm)	R (1/10 mm)	C+300	R+300	Ratio	IQD
0358	1930	15-19 st.	1	98	3092	398	3392	8.5	20.5
6049	1984	23-26 st.	11	0	3678	300	3978	13.3	40.7
8501	1969	27-30 st.	10	0	3309	300	3609	12.0	77.1
9195U	1986	34-38 st.	1	80	8164	380	8464	22.3	4.4
9815E	1979	74-81 st.	1	0	3530	300	3830	12.8	5.5

3.5. Homogeneity control procedure.

In order to detect inhomogeneities in the series, many statistical tests have been developed, reviewed in Szalai et al., (1999), Peterson et al., (1998) and Lanzante (1996). We can point out some of them such as Rhoades and Salinger (1993, test CUSUM), Rodriguez et al., (1999, test of von Neuman), Tayanç et al.., (1998, test of Kruskal-Wallis), Easterling and Peterson (1995, likelihood ratio), Gan (1995, test of Kendall), Lanzante (1996, test of Wilkoxon-Mann-Withney), Tarhule and Woo (1998, tests of Pettit, Man Withney and Man Kendall), and of course SNHT test of Alexanderson (Alexanderson, 1986; an excellent overview of SNHT can be found in Keiser and Griffits, 1997).

On the other hand, many meteorological services developed their own methods (see specific cases, Vincent and Gullet, 1999; Tuomenvirta, 2001; Peterson et al., 1998), whose software is often not available to the public. However, the slant of using suitable test depends on the software availability as well as on the contacts with the developers.
In our case, the control of homogeneity in the depurated data base has been developed by SNHT (Alexanderson, 1986), both for monthly and seasonally scale. The softwares used in this analysis are AnClim and ProClimDB (Štepánek, 2005a, 2005b).

In order to select the statistical inhomogeneities that had to be corrected, we used temporary windows, supported by metadata obtained during the detection process of the series with repeated data. In any case, when inhomogeneities have been detected, those were corrected in complete years. We didn't accept inhomogeneities neither in the initial nor in the final part of the series (10 years). Also we only accept and correct detected inhomogeneities affecting at least at 3 independent months in the same period (using a window of 5 years). After carrying out the corrections, we proceeded to calculate new reference series and applied again the SNHT test. Total of considered inhomogeneous series was of 202 in both applied runs.

Finally, to verify the effect of depuration process on the homogeneity tests, we also applied SNHT test on the original data and the first reference series (obtained before the first step of anomalous data detection).

3.6. The reconstruction processes.

The final step consisted of making reconstruction of all depurated and homogenous series by means of new reference series.

For each observatory, two reference series were calculated by using the same procedure described before, in this case, with maximum distances of 10km and 25km (Reference series Nº 12 and 13) between neighbours. The series were previously manipulated in order to make it possible to join two closed series with successive period of time. Otherwise, we were not able to obtain prolonged series by combining series of successive period of time. This fact affected particularly all denominated secular or historical series begins before 1900.

The final filling up of gaps was directly made with data of the reference series obtained at 10 km and 25 km.

4. RESULTS

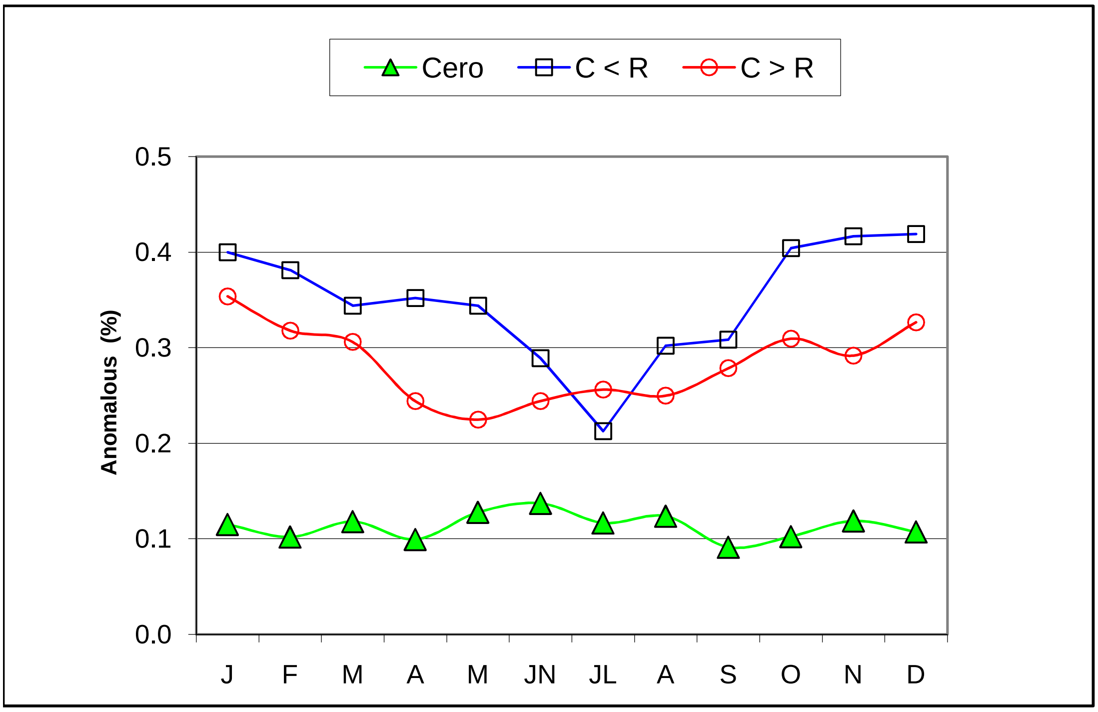


Fig. 2. Monthly distribution of anomalous data. (C candidate, R Reference)

The total number of considered anomalous data finally discarded was 7182 (approximately 0.75 %), whose monthly distribution is shown in Figure 2. The data are classified as positive anomaly (i.e. C>R), negative (C<R) and zero anomalous. Zeros don't show monthly pattern, while high and low suspicious data occures mainly in winter (Figure 2).

These data are temporally accumulated in two periods: around 1925 and 1960. The smallest amount of data can be found around the period of civil war and the final decade (Figure 3). However, as the number of operative observatories and those of registered data are very variable in time, these values must be weighted.

In Figure 4, we show the evolution of the number of anomalous data in proportion to data registered in each year. We clearly appreciate that the density of anomalous data is located in the first half of the 20th century and decreased considerably from 1950, even though it is the period of greater number of operating observatories and registered data. The data correspond only to the period of 1901-2000, for it was not possible to contrast with reference series of time period previous to 1900, although the total contrasted data with its respective reference ascend to 99 %.

In the original series, without depurating process, 2984 statistical inhomogeneities were detected, and they affected 1966 series of precipitation, which suppose to be 75% of the

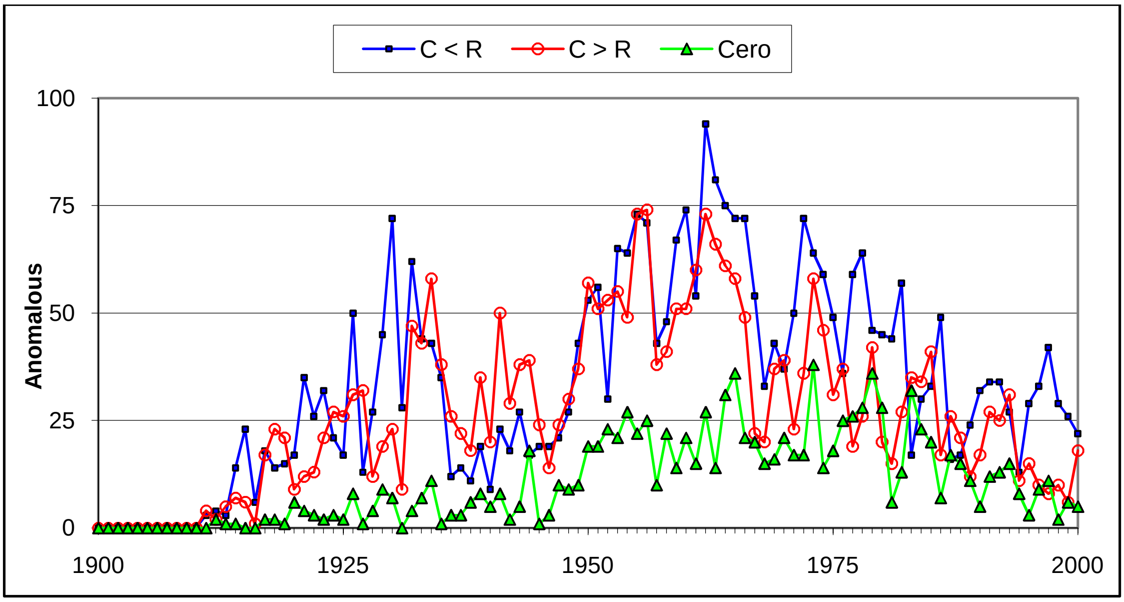


Fig. 3. Annual evolution of anomalous data. (C candidate, R Reference)

total series. Nevertheless, in the depurated data, only 1125 inhomogenous series were detected (43%), including 1795 statistical inhomogeneities. The results of this analysis are shown in Figure 5, where it is compared with the evolution of the total observatories operating in each year. In the first graph (upper), we show the total number of inhomogeneities, with and without depuration, detected in each year. In the second one (low), we show the annual index of inhomogeneity to avoid the effect of the number of observatories. By analyzing both graphs, we can see that no depurated series presents greater number of inhomogeneities over time, and they decrease from the second half of the 20th century. Finally the total of considered inhomogenous series was of 202.

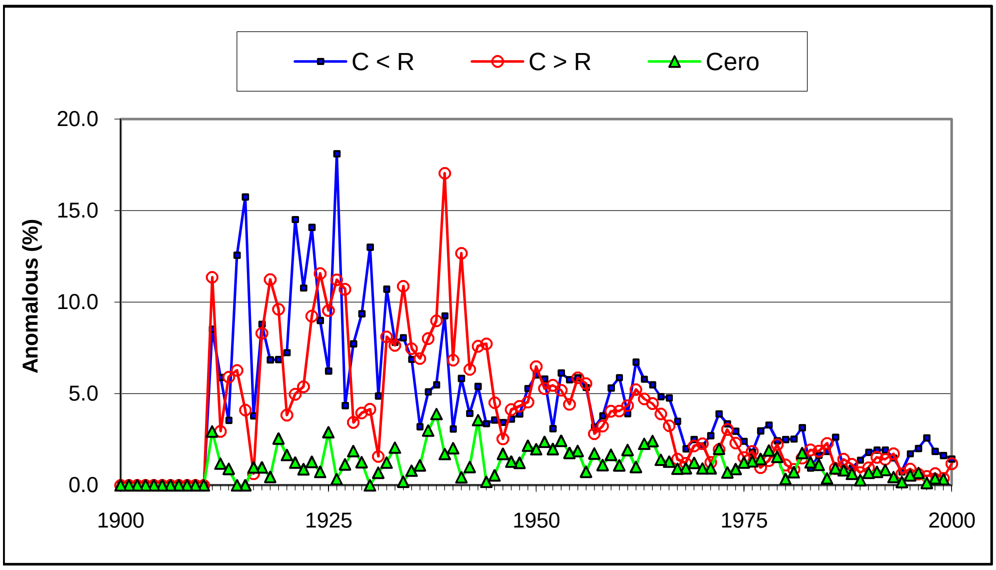
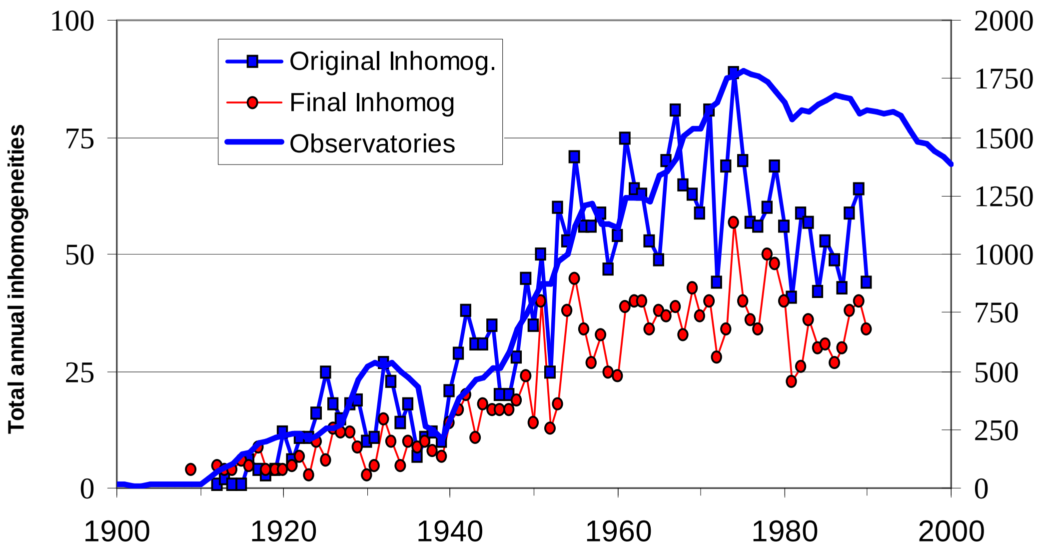


Fig. 4. Annual index of anomalous data. (C candidate, R Reference)



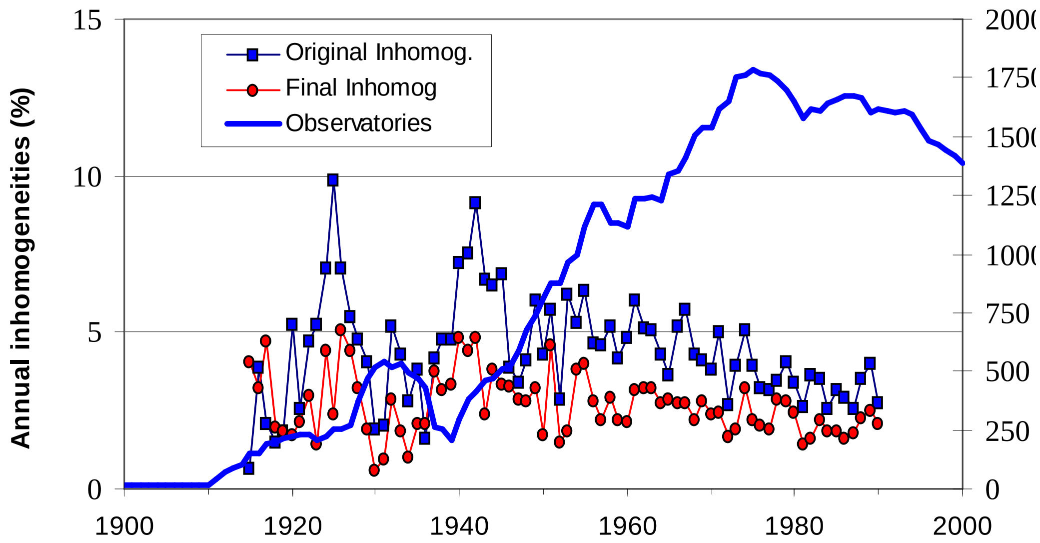


Fig. 5. The effect of anomalous data detection on inhomogeneity analysis.

The obtained data base, regarding its special density, begins practically in 1915 and the number of observatories increases until the Spanish Civil War period (1936-1939). During that time, a drop shows up, but since then, the number of observatories increases again until the 70’s, when it reaches its maximum number. Then, it descends until now in many places, may be in consequence of the rural exodus (Figure 6).

This data base is considered as a result of the reconstruction of the total analyzed and homogenized series (2669), therefore, we must be careful when trying to use it. In fact, combining criteria of minimum length of original data, few numbers of gaps, etc, for the second half of the 20th century, we considered about 1113 observatories which fulfill all the previous criteria. Thus, they can be used to analyze the evolution of precipitation in the Eastern part of the Iberian Peninsula. In Figure 7, we show the spatial distribution of the data base, although there are areas with low density.

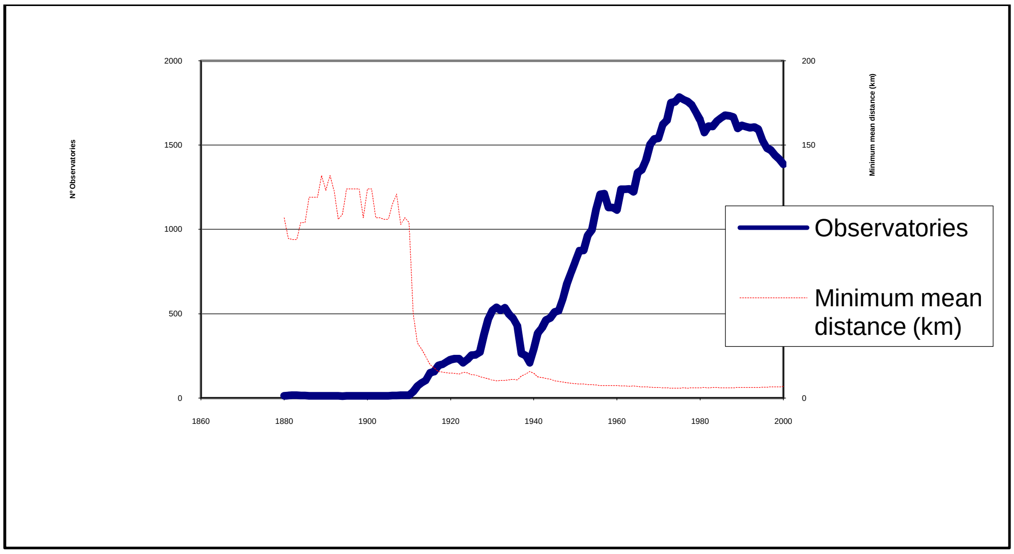


Fig. 6. Number of observatories and Minimum mean distance in km

Problem of areas over 1500 m o.s.l. persists, about which the information is still lacking.

Fig. 7. Spatial distribution of data base.

5. SOME INITIAL RESULTS.

The descriptive analysis of precipitation trend during winter months (December-March) from 1951-2000 is shown in a collection of maps (Figures 8-9-10-11). All of them show significant (positive (+) blue, negative (-) red) and not significant trends (n.s. dot).

Trends were calculated at p 0.1 level, after low pass filter (9 laps) by using Spearman rank order correlation. The results show a very clear spatial pattern from coastland-inland and North-South along winter months, and are very promising for future analysis.

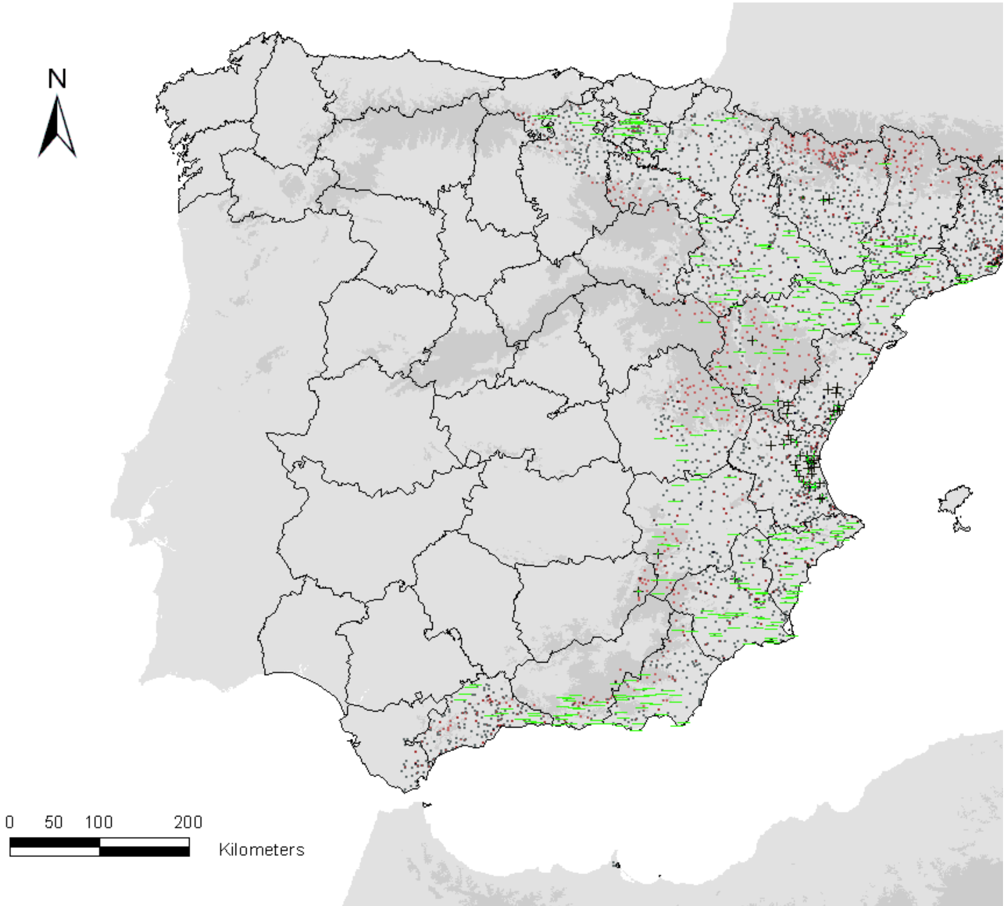


Fig. 8. December precipitation trend (1951-2000)

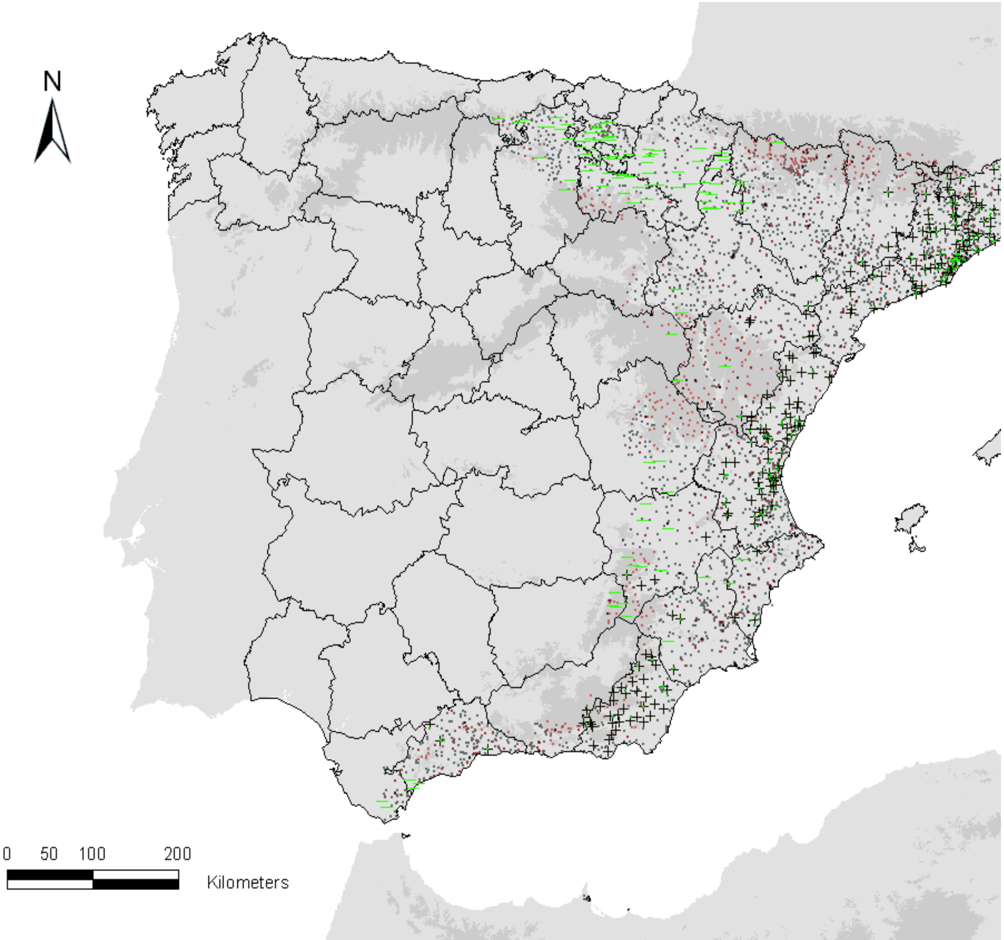


Fig. 9. January precipitation trend (1951-2000)

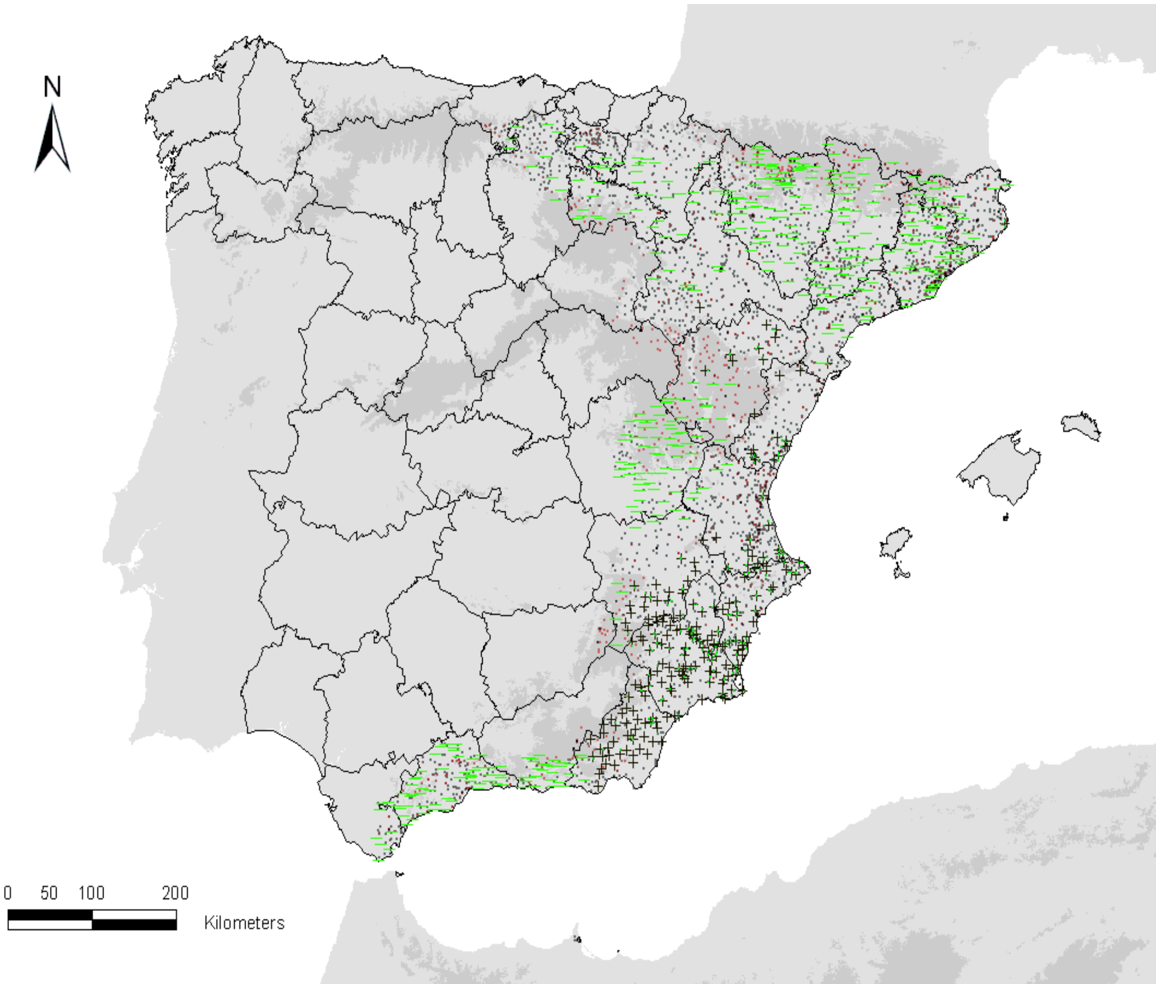


Fig. 10. February precipitation trend (1951-2000)

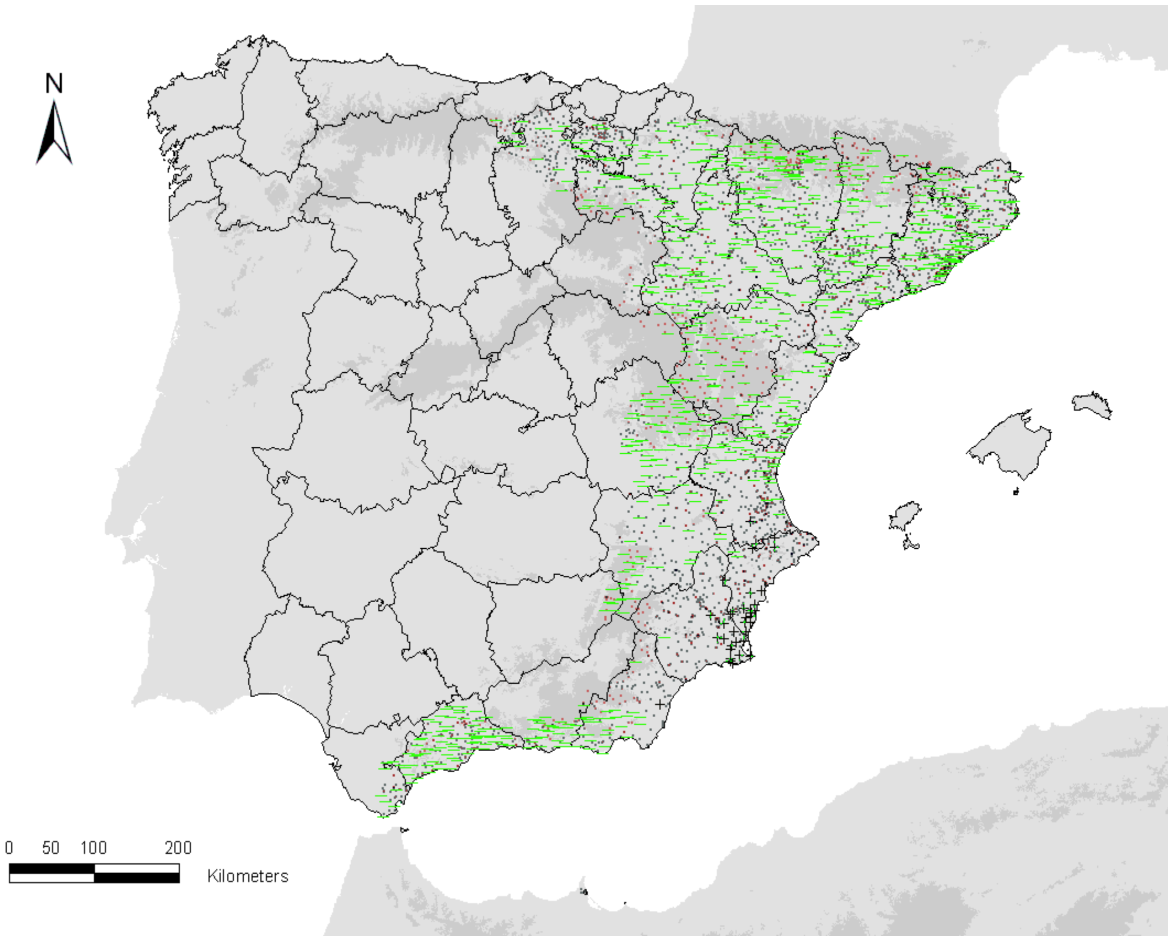


Fig. 11. March precipitation trend (1951-2000)

CONCLUSIONS

Our analysis shows that detection and correction of suspicious data in a precipitation database seems to be preliminary and necessary task to avoid many statistical inhomogeneities.

Following this procedure we have built up a dense data base which covers more than 1/3 of the Iberian Peninsula with spatial density circa (?) 1 observatory / 150 km^2. The high density of our data base enables us to do in the future spatial analysis of sub-regional models and their empirical validation with *in situ* measurements. The provisional result of monthly trends shows a very coherent spatial pattern and involves that not great mistake or bias calculations have been produced.

There is a lack of data in areas >1500 m o.s.l. It would be interesting to seek for suitable methodologies, with the aim of making it possible to fill data up spatially, in order to be able to use them for regional climate models.

ACKNOWLEDGEMENTS

We thank the Ministry of Science and Technology of Spain (projects CGL2005-04270, REN2003-07453, REN2002-01023-CLI) for his financial support. Thanks also go to the General Government of Aragón (Consolidated Research Group of Climate, Water, GlobalChange and Natural Systems, Resolution 13 of April 2005, BOA 20-04-2005).

REFERENCES.

Alexanderson, H. (1985): A homogeneity test applied to precipitation data. Journal of Climatology, 6, 661-675.

Allen, M.R. and Ingram, W.J. (2002): Constraints on future changes in climate and the hydrological cycle. Nature, 419, 224-232.

Balairón L. (2000): Las causas del cambio climático. En L. Balairón (Coord.) El Cambio climático. Servicio de Estudios del BBVA, 137, 89-109.

Barrow E., Hulme M. and Semenov M. (1996): Effect of using different methods in the construction of climate change scenarios: Examples from Europe. Climate Research, 7, 195-211.

Cosgrove C.M. and Garstang M. (1995): Simulation of rain event from raingauge measurements. International Journal of Climatology, 15, 1021-1029.

Easterling D. and Peterson T. (1995): A new method for detecting undocumented discontinuities in climatological time series. International Journal of Climatology, 15: 369-377.

Esteban-Parra, M.J., Rodrigo, F. and Castro-Díez, Y. (1998) Spatial and temporal patterns of precipitation in Spain for the period 1880-1992. International Journal of Climatology, 18, 1557-1574.

Gan T.Y. (1995): Trends in air temperature and precipitation for Canada and north-eastern USA. International Journal of Climatology, 15, 1115-1134.

González Hidalgo, JC, de Luís M., Vicente, SM, Štepánek, P., Raventós, P., Saz. MA, Cuadrat, JM, Ferraz, J. and Creus, J. (2004) Monthly precipitation data base in mediterranean climate area of Spain. Reconstruction processes and quality control.

World Meteorological Organization, WCDMP-No. 56, WMO-TD No. 1236, 105-116.
González Rouco J.F., Jiménez J.L., Quesada V. and Valero F. (2001): Quality control and homogeneity of precipitation data in Southwest of Europe. International Journal of Climatology, 14, 964-978.
Groisman P.Ya. and Legates D.R. (1994): The accuracy of United States precipitation data. Bulletin American Meteorological Society, 75, 215-227.
Houghton J.T., Ding Y., Griggs D.J., Noguer M., van der Linden P.J. and Xiaosu D. (Eds) (2001): Climate change 2001. The scientific basis. Cambridge University Press, UK, 896 pp.
Hulme M. (1995): Estimating global changes in precipitation. Weather, 50, 34-42.
Hulme M., Conway D., Jones P.D., Jiang T., Barrow E.M. and Turney C. (1995): Construction of a 1961-1990 European climatology for climate change modeling and impact applications. International Journal of Climatology, 15, 1333-1363.
Hurrel J.W. (1995): Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. Nature, 269, 676-679.
Jones P.D. and Conway D. (1997): Precipitation in the British isles: an analysis of area-average data updates to 1995. International Journal of Climatology, 17, 427-438.
Jones P.D. and Hulme H. (1996): Calculating regional climate time series for temperature and precipitation: methods and illustrations. International Journal of Climatology, 16, 361-377.
Keiser D. and Griffiths J. (1997): Problems associated with homogeneity testing in climate variations study: a case study of temperature in the northern Great Plains USA. International Journal of Climatology, 17: 497-510.
Lana, X., Martínez M.D., Serra, C. and Burgueño, A. (2004): Spatial and temporal variability of the daily rainfall regime in Catalonia (NE Spain), 1950-2000. International Journal of Climatology, 24, 613-641.
Lanzante L.R. (1996): Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples including applications to historical radiosonde station data. International Journal of Climatology, 16, 1197-1226.
Martín-Vide, J. (2001): El índice NAO y la precipitación mensual en la España peninsular. Investigaciones Geográficas. 26, 41-58.
Matyasovszky I., Weidinger T., Bartholy J. and Barcza Z. (1999): Current regional climate change studies in Hungary: a review. Geographia Helvetica, 54 (3), 138-146.
Mcnamara P. (1999): Global Changes, Regional Impacts: Climate Change In The Midle East. Geographia Helvetica, 54 (3), 132-137.
Mearns L.O., Giorgi F., McDaniel L. and Shields C. (1995): Analysis of daily variability of precipitation in a nested regional climate model: comparison with observations and doubled CO2 results. Global Planetary Change, 10, 55-78
Moron V., Bigot S. and Roucou P. (1995): Rainfall variability in subequatorial America and Africa and relationships with the main sea-surface temperature modes (1951-1990). International Journal of Climatology, 15, 1297-1322.
New M., Todd M., Hulme M. and Jones, P. (2002): Precipitation measurements and trends in the twentieth century. International Journal of Climatology, 21, 1899-1922.
Palutikof J.P., Goodess C.M. and Hacot H. (1996): Future rainfall trends over the mediterranean. Book of abstracts. International conference on mediterranean desertification. Crete-Hellas, 29 Oct - 1 Nov 1996. Abstract book, (no pages).
Parry M. (2001): Climatic change: where should our research priorities be?. Global Environmental Change, 11, 4, 257-260.

Pérez Cueva A. (Dir). (1994): Atlas climàtic de la Comunitat Valenciana. Direc. General d'Urbanisme i Ordenació del Territori, Generalitat Valenciana, Col.lecció Territori, Nº 4.
Peterson T. and Easterling D. (1994): Creation of homogeneous composite climatological reference series. International Journal of Climatology, 14, 671-679.
Peterson T., Easterling D., Karl, T. et al. (1998): Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18, 1493-1517.
Prudhomme C., Reynard N. and Crooks S. (2002): Downscaling of global climate models for flood frequency analysis: where are we now? Hydrological processes, 16, 1137-1150.
Rhoades D. and Salinger M. (1993): Adjustment of temperature and rainfall records for site change. International Journal of Climatology, 13, 899-913.
Rodríguez R., Llasat M.C. and Wheeler D. (1999): Analysis of the Barcelona precipitation series 1850-1991. International Journal of Climatology, 19, 787-801.
Rodríguez-Puebla, C., Encinas, A., Nieto, S. and Garmendia, J. (1998): Spatial and temporal patterns of annual precipitation variability over Iberian Peninsula. International Journal of Climatology, 18, 229-316.
Rodrigo F.S., Esteban Parra M.J., Pozo D. and Castro-Díez Y. (2000): Rainfall variability in southern Spain on decadal to centennial times scales. International Journal of Climatology, 20, 721-732.
Szalai A., Szentimrey T. and Szinell C. (1999): Proceedings of the second seminar for homogeneization of surface climatological data. WMO-TD Nº 962, Géneva.
Štěpánek P. (2005 a): AnClim - software for time series analysis. Dept. of Geography, Fac. of Natural Sciences, MU, Brno, 1.47 MB.
http://www.sci.muni.cz/~pest/AnClim.html.
Štěpánek, P. (2005 b): ProClimDB – software for processing climatological datasets. CHMI, regional office Brno. http://www.sci.muni.cz/~pest/ProcData.html
Sulzman E.W., Poiani K.A. and Kittel T.G.F. (1995): Modeling human-induced climatic change: A summary for environmental managers. Environmental Management, 19, 197-224.
Sumner, G, Ramis, C and Guijarro, J.A. (1993): The spatial organization of daily rainfall over Mallorca, Spain. International Journal of Climatology, 13, 89-109.
Tank, A.M., Wijngaard, J.B., Können, G. et al. (2002): Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessement. International Journal of Climatology, 22, 1441-1453.
Tarhule A. and Woo M. (1998): Changes in rainfall characteristics in northern Nigeria. International Journal of Climatology, 18, 1261-1271.
Tayanç M., Dalfes H., Karaca M. and Yenigün O. (1998): A comparative assessment of different methods for detecting inhomogeneities in Turkish temperature data set. International Journal of Climatology, 18, 561-578.
Tuomenvirta, H. (2001): Homogeneity adjustments of temperature and precipitation series-Finnish and Nordic data. International Journal of Climatology, 21, 495-506.
Vincent L. and Gullet D. (1999): Canadian historical and homogeneous temperature datasets for climate changes analyses. International Journal of Climatology, 9, 1375-1388.
Vinnikov K.Ya., Groisman P.Ya. and Lugina K.M. (1990): Empirical data on contemporary global climate changes (temperature and precipitation). Journal of Climate, 3, 662-677

Wilby R.L. and Wigley T.M. (2000): Precipitation predictors for downscaling observed and general circulation models relationships. International Journal of Climatology, 20, 641-661.

Zorita E. and Gonzalez Rouco F. (1999): Divergencias de las predicciones de la futura intensidad de la oscilación el Atlántico Norte. En J. Raso y Martín Vide, J. (Eds) La climatología española en los albores del siglo XXI. Publicaciones de la Sociedad Española de Climatología, 1, 579-587.

HOMOGENIZATION AND VALIDITY CONTROLS FOR TEMPERATURE TREND ESTIMATES OVER ITALY

Andrea Toreti and Franco Desiato
APAT, Via Vitaliano Brancati 48, 00144 Rome – Italy
tel. +390650072567, fax +390650072657
andrea.toreti@apat.it, franco.desiato@apat.it

INTRODUCTION

In order to give an answer to the needs of harmonisation and standardisation of climate indicators calculation, and of a fast and reliable update and access to the data, a system denominated SCIA (Sistema nazionale per la raccolta, elaborazione e diffusione di dati climatologici di interesse ambientale) was realised by the Italian environmental protection agency (APAT), in collaboration with the main national and regional meteorological institutions. In this framework, the processing of meteorological data coming from the synoptic network of the Italian Air Force weather service is included. Data recorded from 49 synoptic stations, characterized by completeness, continuity and good geographical distribution (fig.1), were selected for time series homogenization and testing of homogenization procedures. We started up this activity by considering the mean monthly temperature time series, derived from daily mean temperature, calculated as the arithmetic average of maximum and minimum daily temperature reported in SYREP messages.

Our aim is to obtain a reliable estimation of the temperature behaviour over Italy in the last decades, filtering the non-climatic factors like relocation and instrument changes.

Fig. 1 Geographical distribution of the 49 stations.

METHODOLOGY

Input data can be affected by errors for several, different reasons. In order to filter evidently wrong data at the origin, we firstly apply a so-called weak climatological control to maximum and minimum daily temperatures included in SYREP message: their values must fall within a range of physically admitted values. This control allows to identify (and reject) gross mistakes, such as typing errors. Data validated through this first quality control, contribute to the calculation of monthly mean values. A second quality control is performed to identify outliers of monthly means. Following the idea of Eisched et al. (1995), the data are clustered as a function of season, and the latitude and altitude of the station, hence outliers dropping outside a range of values which is a function of the interquartile range of each cluster, are found (Baffo et al. 2005). Outliers are then checked through the analysis of input data in terms of time continuity and space correlation with data of nearby stations. If it is confirmed that a data is wrong and cannot be corrected, it is rejected. If the number of missing or non-valid data in a month exceeds a threshold of 25%, the monthly value is eliminated from the series. Finally we have a validated monthly series to test and eventually homogenize.

After an analysis of the available statistical procedures for homogenization (Aguilar et al., 2003; Peterson et al., 1998), we chose a parametric one, i.e. the Standard Normal Homogeneity Test - SNHT developed by Alexandersson (Alexandersson, 1986; Alexandersson and Moberg, 1997). It is important to mention some choices we made in the test application. First of all we decided to use the single shift version of the test; then we have calculated correlation coefficients (between the so-called candidate station and the others) using the transformed first difference series, as highlighted by Peterson and Easterling (1994). The best correlation criterion, with some geographical limitations, indicates that five stations (the recommendation is at least three) applicable for the creation of the reference series; we want to emphasize the importance of this step, because the reference series should reveal the climatic behaviour and strongly influences the following results.

The practical application of SNHT involves a first test of the entire series; if a shift is detected the series is divided into two periods that are tested separately; if one more shift is detected, this procedure is repeated, until a homogenous or a too short period is found. The correction of inhomogeneous time series has been done following the philosophy of 'reproducing current measure conditions'; therefore the oldest periods have been 'adapted' to the most recent. In the case there is more than one shift. We corrected the series from the most recent shift forward (until 2004), reapplied the test to the entire series and corrected the oldest shift.

RESULTS

In this section we show the results of the test-homogenization session. Two practical examples are used to better understand the principal features of the statistical tool and the differences between annual mean temperature anomalies over Italy calculated before and after homogenization.

Among the 49 selected monthly time series, we found 6 homogeneous series, 43 series with at least one inhomogeneity and 28 series with more than two inhomogeneities. The initial year of the final series (homogenized and homogeneous) is determined by two factors: the first year of station recording and the availability of three stations for the reference series (fig.2). 1951 and 1961 are the years with the largest number of time series starts.

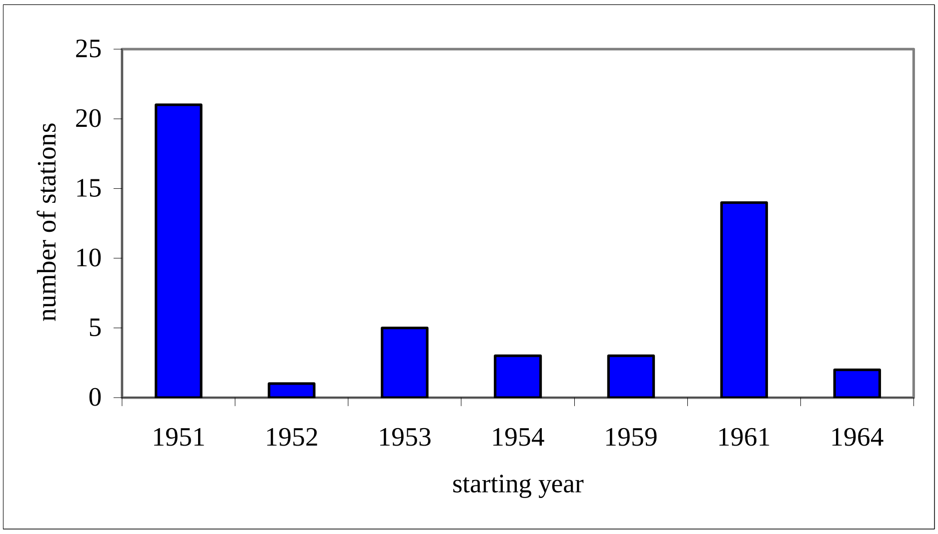


Fig. 2. Number of stations in function of the starting year.

There are many stations with more than one inhomogeneity, and the most common number is three (Fig.3).

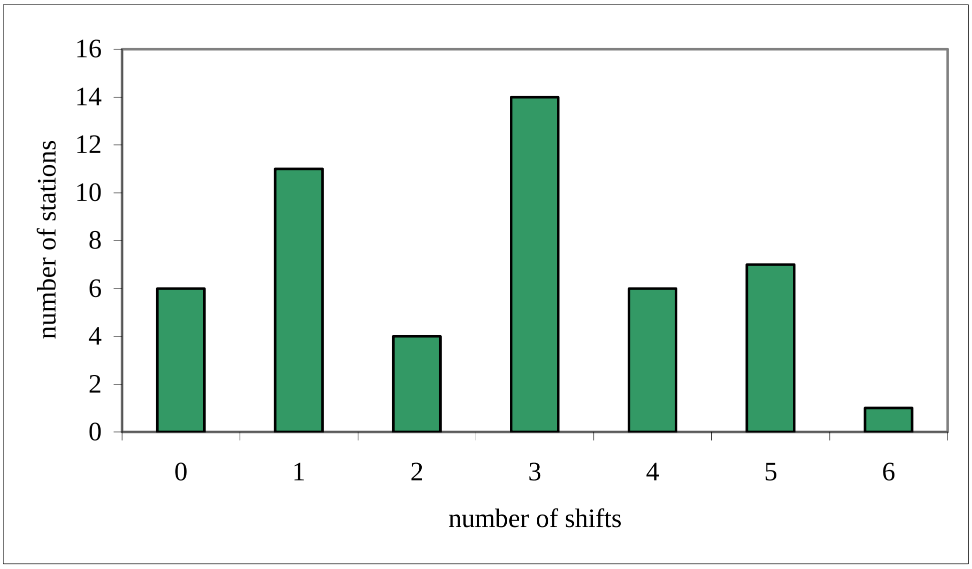


Fig. 3. Number of stations in function of the number of shifts.

The analysis of the number of shifts as function of the year of occurrence shows that there is not a change that involves the entire network in a specific year (fig.4).
There are only four years with no supposed changes (1952, 1980, 1996, 1997), since the last and the first year cannot be considered.

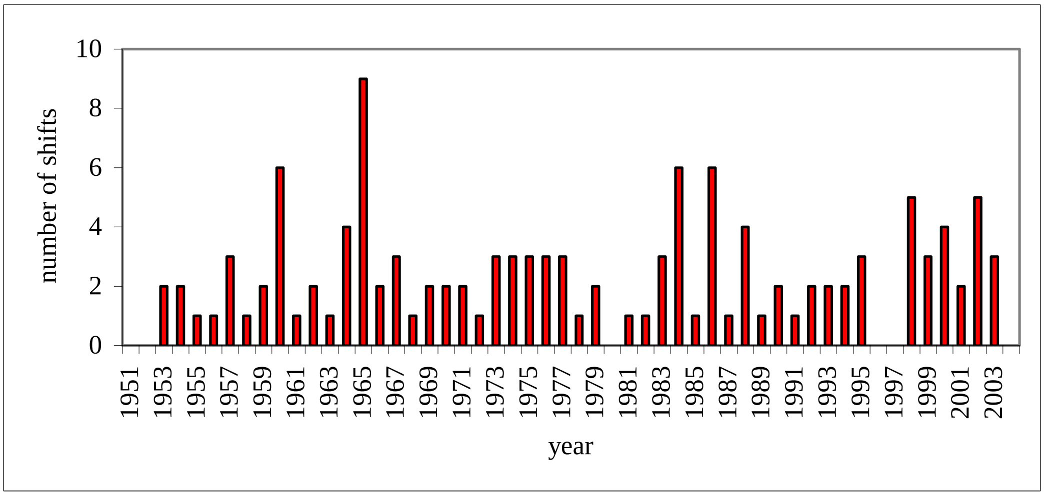


Fig. 4. Number of shifts in function of the year.

Finally, the shifts are not concentrated in specific months, having all months at least five shifts, with the largest occurrence in March and September (Fig.5).

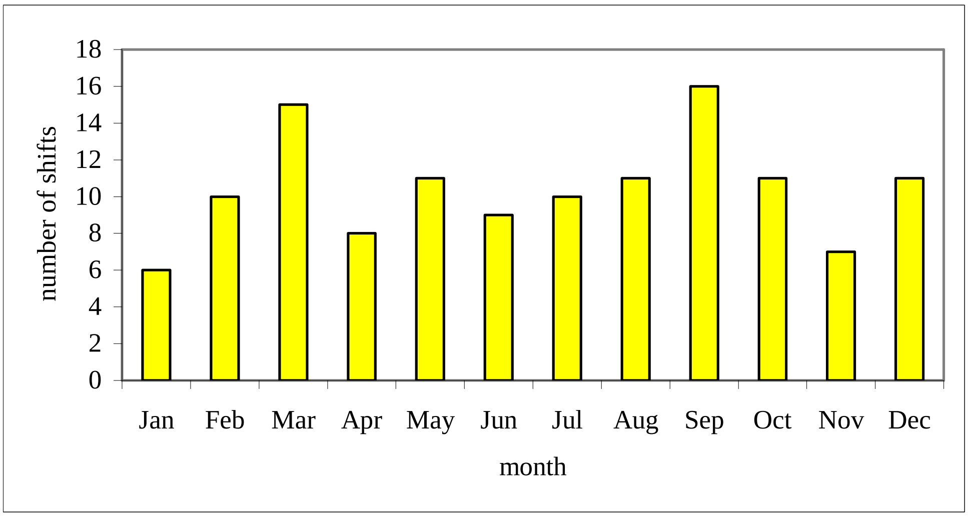


Fig. 5. Number of shifts in function of month

The first practical example, that we want to show, is the homogenization test of the Rome Ciampino station (WMO code: 16239). The reference series has been constructed with five stations (16234, 16244, 16224, 16243, and 16245) that have a correlation coefficient greater than 0.985; we have found three shifts in the following month/year: February/1964; March/1973; August/1977. The corrections were applied and the original and homogenized monthly series are presented in fig.6. The differences between the two monthly series are hidden by the seasonal cycle, while they are evident in the annual series (Fig.7).

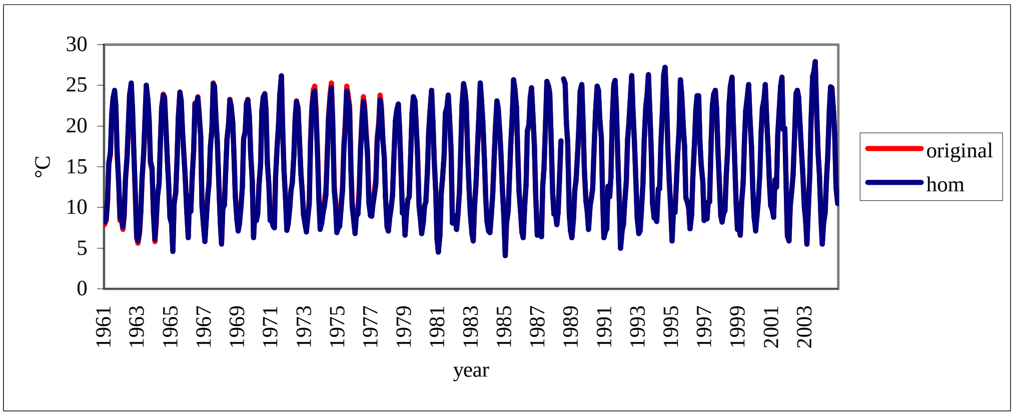


Fig. 6. Monthly series of Rome/Ciampino, homogenized (blue) and original (red).

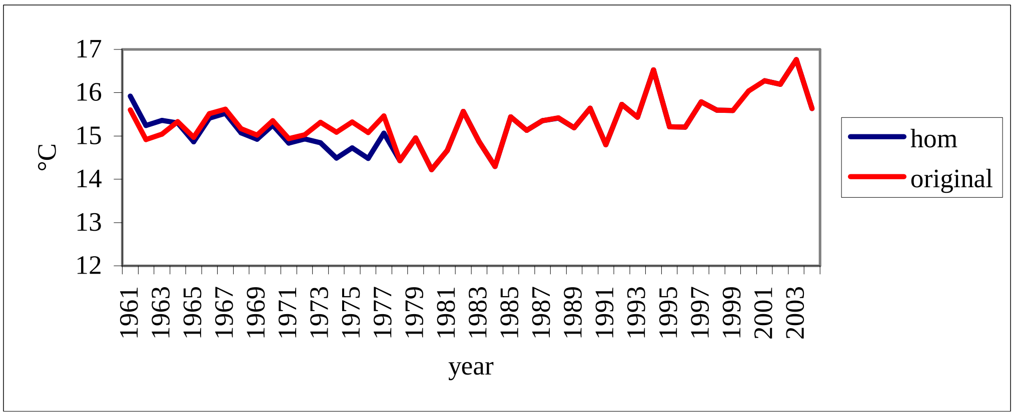


Fig. 7. Annual series of Rome/Ciampino, homogenized (blue) and original (red).

Another station (16206), besides the five chosen for the reference series, has a high correlation coefficient, but presents some problems in terms of homogeneity. Then, we tested the sensitivity of the homogenization procedure with respect to the composition of the group of stations used to calculate the reference series. Four tests were carried out: two with five stations and two with three stations; in both cases the station 16206 is firstly included and then excluded from the group. Available metadata document indicates a relocation of the Rome Ciampino station in 1977; so we decided to begin the analysis after the shift of 1964 and to test the ability of the SNHT to detect the shift occurred in 1977.
The t statistic of the 'five stations test' gives good results both with and without station 16206 (fig.8), i.e. the 1977 shift is well identified; the 'three stations test' shows a different behaviour of t in the two cases (fig.9). When the 16206 station is excluded the statistic is still able to detect the 1977 shift, and the entire behaviour of the function is similar to that of the 'five stations test'; the introduction of 16206, which is inhomogeneous itself, compromises the t series.

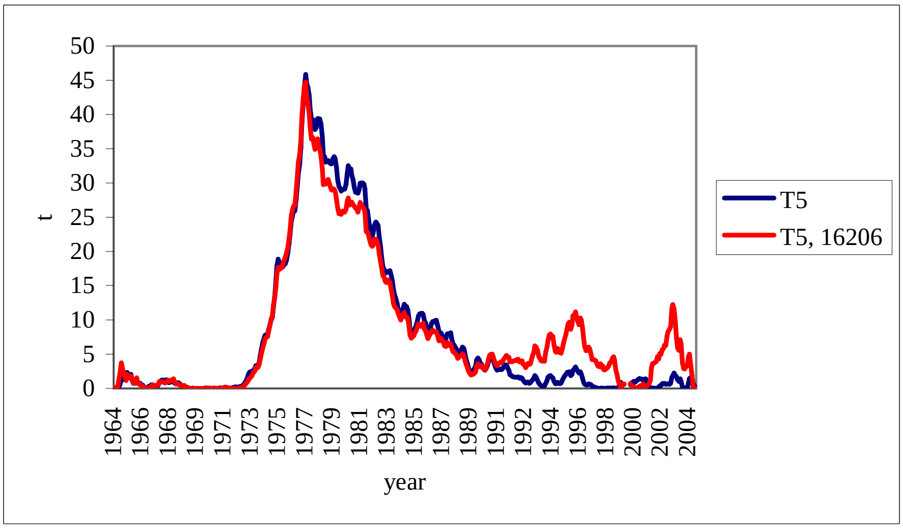


Fig. 8. T statistic with five stations. Without (blue) and with (red) station 16206.

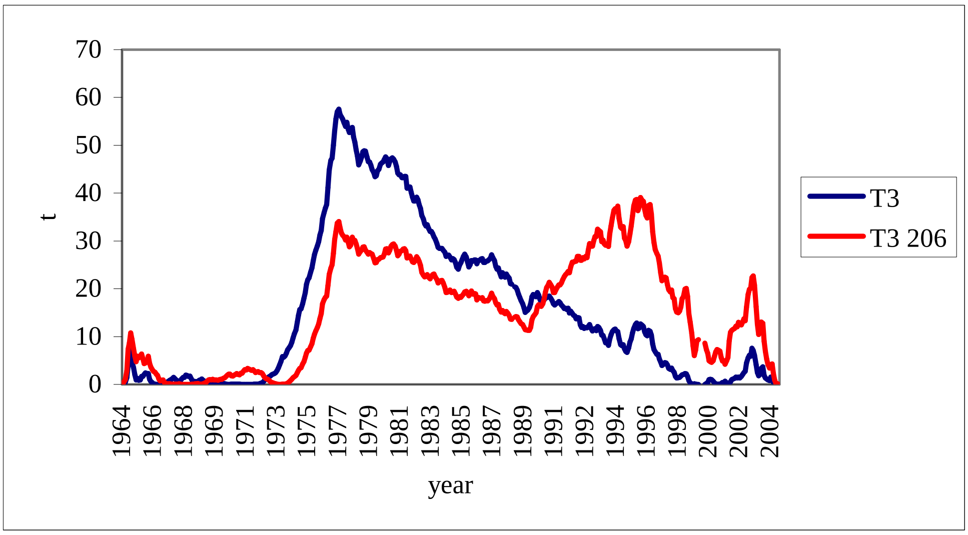


Fig. 9. T statistic with three stations. With (red) and without (blue) 16206.

This test makes it appearthat the number of stations and a careful selection of the stations to be used for the calculation of the reference series, may be very important, and that a wrong choice may lead to unrealistic results.

The second example involves the station Foggia/Amendola (WMO code 16261), with the reference series obtained using five stations (16312, 16232, 16320, 16332, 16360). This series has only one shift in April 1965, so the correction procedure is fast and easy; fig.10 compares the annual homogenized series to the original one, while fig.11 shows the statistic before and after the homogenization.

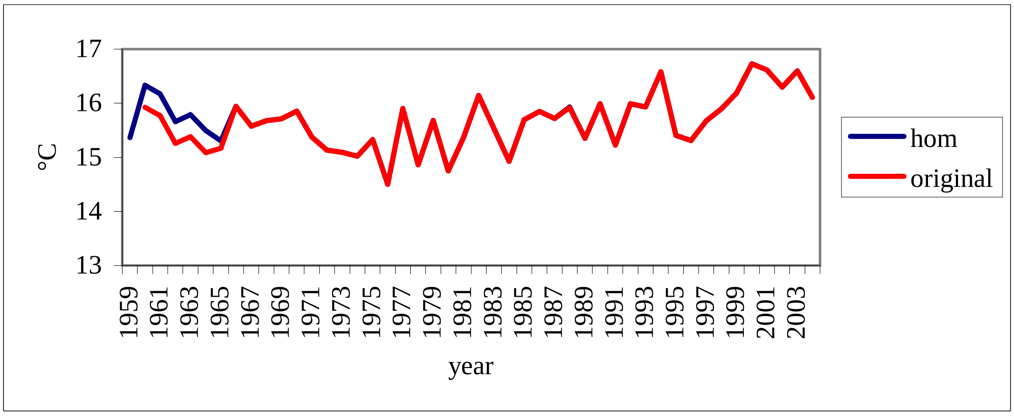


Fig. 10. Annual series of station 16261, homogenized (blue) and original (red).

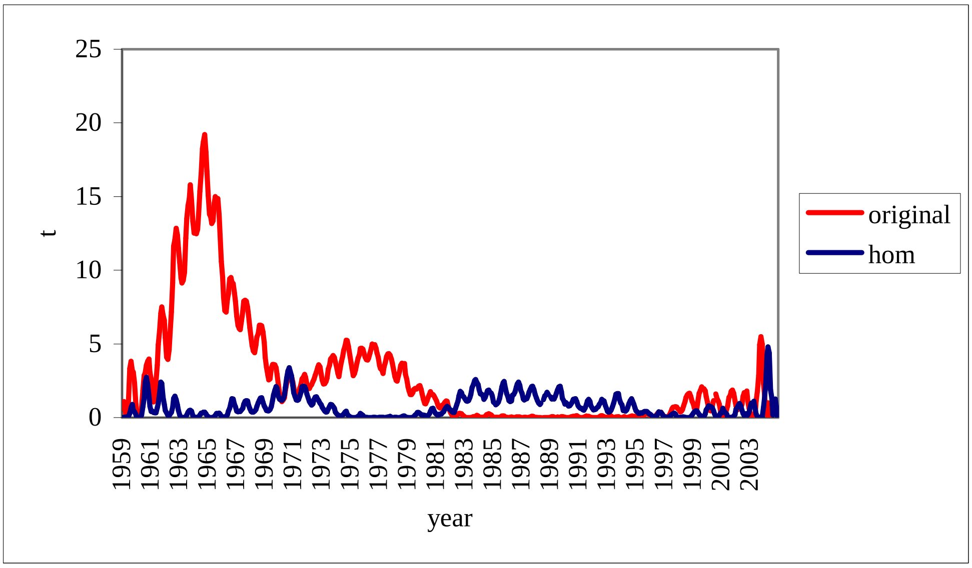


Fig. 11. T statistic of 16261, original (red) and homogenized (blue).

Finally, the comparison of the annual mean temperature anomaly over Italy from 1961 until 2004 (obtained by averaging the anomaly values over all 49 stations), calculated before and after homogenization, is presented. Fig.12 shows the values of differences, ranging form –0.06 to 0.07 °C. This relatively small effect of homogenization on the mean anomaly series may be due to several reasons: the different number, sign, amount and year of occurrence of the inhomogeneities; the fact that the network has to respect WMO technical specifications, that limit the possible changes of stations characteristics; the fact that these stations are mainly located in airport areas, where altering factors such as growing heat island or large relocations are reduced or improbable.

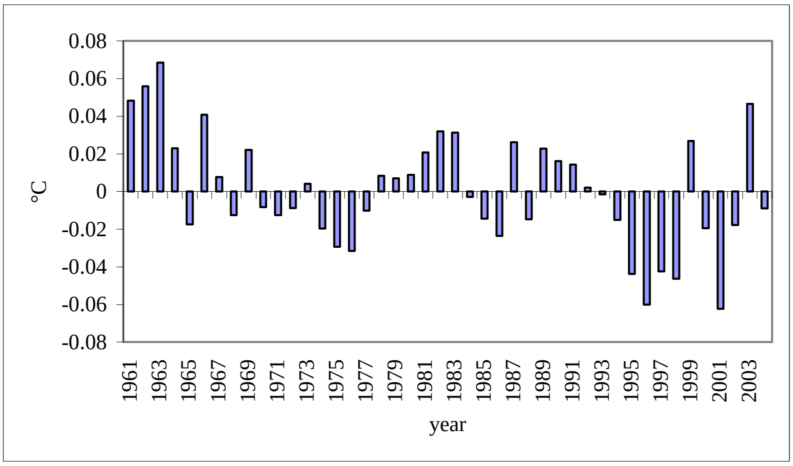


Fig. 12. Difference series between the homogenized annual mean temperature anomaly and the original series.

CONCLUSIONS

In the context of SCIA project 49 monthly time series coming from the network of the Italian Air Force Weather Service have been tested and homogenized, in order to obtain a reliable estimation of annual mean temperature anomaly over Italy. Six series came out to be homogeneous, while twenty-eight have more than two inhomogeneities. The shifts do not occur in a specific month/year or period, and also the sign and the amount of the shifts are very different among the series. Two examples of homogenization test were presented. The first gave us the possibility to show the sensitivity of SNHT with respect to the number and the choice of the stations used to calculate the reference series. The results indicate that the use of five series is advisable. Finally, the differences between annual mean temperature anomaly coming from the homogenized set and the original one have been shown.

REFERENCES

Aguilar, E, I. Auer, M. Brunet, T.C. Peterson, J. Wieringa, 2003: Guidelines on climate metadata and homogenization. WMO/TD No.1186, WCDMP 53.

Alexandersson, H, 1986: A homogeneity test applied to precipitation data. J. of Climatol., 6, 661-675.

Alexandersson, H, A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. Int. J. of Climatol., 17, 25-34.

Baffo, F, B. Suatoni, F. Desiato, 2005: Indicatori climatici: i controlli di validità e la ricerca dei valori errati. Bollettino Geofisico, 1-2, 31-43.

Eisched, JK, C. Bruce Baker, T.R. Karl, H.F. Diaz, 1995: The quality control of long-term climatological data using objective data analysis. J. of Applied Meteorol., 34, 2787-2795.

Peterson, TC, D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series. Int. J. of Climatol., 14, 671-679.

Peterson, TC, D.R. Easterling, T.R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Böhm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta,O. Mestre, T. Szentimrey, J. Salinger, E.J. Førland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, D. Parker, 1998 : Homogeneity adjustments of in situ atmospheric climate data: a review. Int. J. of Climatol., 18, 1493-1517.

DETECTION OF INHOMOGENEITIES IN WIND DIRECTION AND SPEED DATA

Predrag Petrović
Republic Hydrometeorological Institute of Serbia, Kneza Višeslava 66, Belgrade, Serbia & Montenegro, tel. +381/11/3537-804
p.petrovic@hidmet.sr.gov.yu

1. PROBLEMS OF PROCESSING WIND (DIRECTION) DATA

Unlike other weather elements, which might be treated as scalars (and thus as a single value), wind data are generally coupled into a pair of values, independent one from another. Therefore, dealing with wind direction and speed data homogeneity is a bit more complicated. In addition, wind direction is usually given as azimouth, an element with limited range of values that reset to minimum of 0 degrees when maximum of 360 degrees is reached (and reversly). Such element disables common mathematical tools for wind direction data processing (Fig. 1).

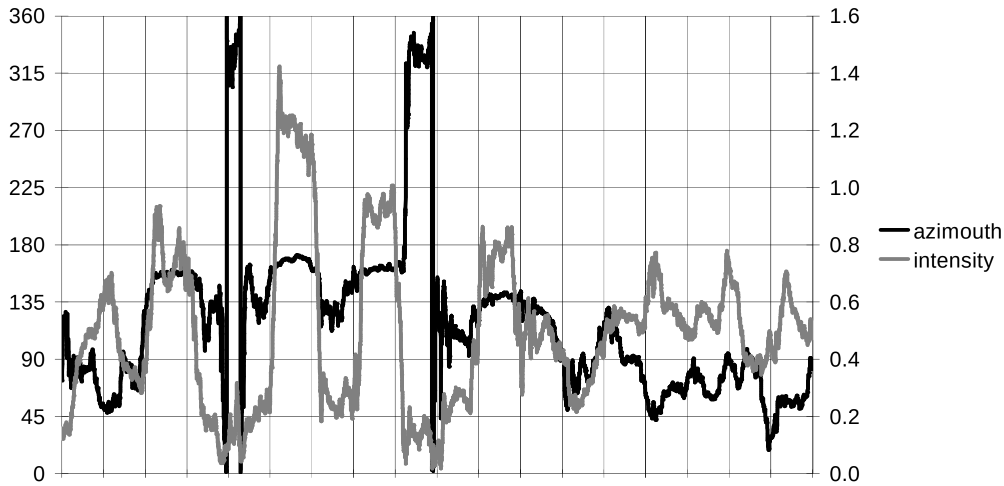


Fig. 1. Annual mean moving values of wind vector azimouth and intensity, Novi Sad - Rimski Šančevi, 1967-1984

Another way to have wind data as scalar value is to convert them into two components. Still, a pair of series with questionable correlation and great fluctuations is not convenient for applying any homogeneity test. Besides, variations in series are of magnitudes that overwhelm possible inhomogeneities (i.e. climate signals, changes in observations / instruments). Thus, splitting of wind vector onto northern and eastern component does not appear to be a good solution of this problem (Fig. 2).

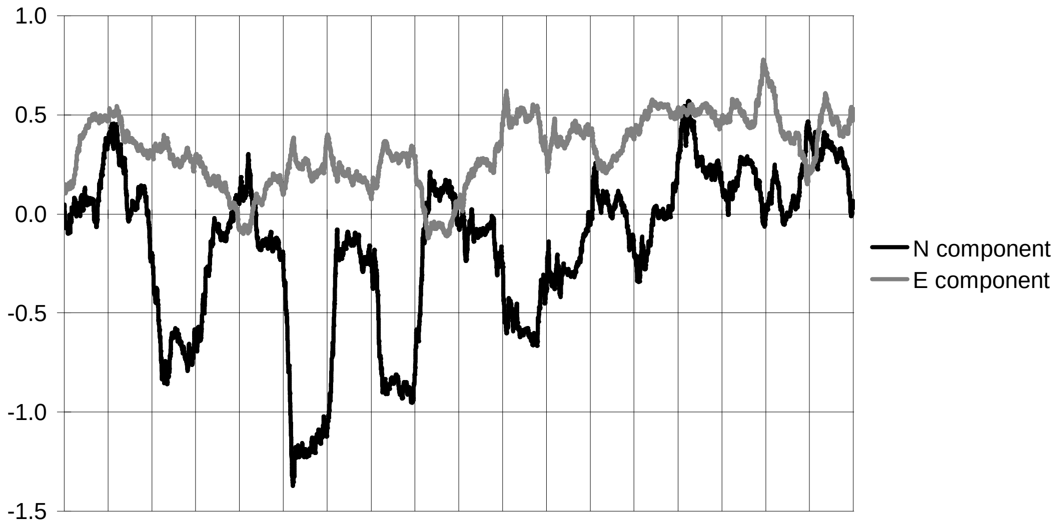


Fig. 2. Annual mean moving values of wind vector north and east component, Novi Sad - Rimski Šančevi, 1967-1984

However, wind direction data in climatology are shown as "wind roses", dealing with distribution of frequences of every distinguished wind direction. Also, some wind speed data processing might include distribution of frequencies of wind speed intervals. Since both direction and speed data might deal with distribution of their values, the ReDistribution Method (Petrović, 2003) might be successfully used for detection of inhomogeneities in these data series.

2. DESCRIPTION OF THE REDISTRIBUTION METHOD

The ReDistribution Method is based on variations in consecutive distributions of frequencies of defined data value classes. In case of wind direction, data value classes are represented by distinguishable wind directions (8, 16, 32 or 36 directions plus calms). The wind speed classes might be wind speed intervals of at least 1 m/s up to Beaufort scale.

Since the processed data window subsets are recommended to be climatologically representative, the length of window subsets that obtain such samples should cover the whole member of years (at least one) and to feature both daytime and nighttime observations (at least two observations per day). In general, using smaller data window subsets might lead to less reliable results that cannot be easily accepted.

The main value is the number of redistributed frequencies between two compared consecutive distributions

$$N_r = \frac{\sum_{i=1}^{n} |d_i|}{2}$$

where d_i is the difference value between two compared frequencies of the same i-th data class (distinguishable wind direction or wind speed interval in this case). Thus, the number of redistributed frequencies is half of the sum of all n absolute differencies d. The ReDistribution Index (*RDI*) is simply the redistributed part of the whole data window subset of N processed data.

$$RDI = \frac{N_r}{N}$$

Theoretically, RDI returns zero where the compared distributions are identical, while RDI eqauls 1 where the compared distributions are entirely different (i.e. from the wind rose that has no south winds to the case of south winds only). In practice, RDI values are always greater than zero because of natural wind fluctuations. Noise level of RDI values depend on climate conditions and the length of data window subset, but it is generally below 0.2. On the other side, RDI value of 1 might be met only in places with seasonal wind roses (i.e. suptropical trade winds) and length of data window that does not cover such seasonal variations).

3. DETECTING BREAK POINTS IN TIME SERIES

Inhomogeneity of data series often includes significant redistribution of values by data classes. The greatest number of redistributed values is reached at the point of the complete change. Further, redistributed values are less featured and returned to the noise level when the new distribution is established. Following this logical principle, peaks in RDI series of wind direction / speed indicate break points in wind data series.

The example of Novi Sad - Rimski Šančevi (Fig. 3) shows a major inhomogeneity break point. Since the RDI peak lies at the beginning of 1982 for the 4-year moving values, homogeneity break dates back to 1978. This peak is featured both in wind direction and wind speed data. This inhomogeneity is confirmed in station metadata. Since this station began to work as a synoptic station, an old wind vane is replaced by anemograph in order to obtain more accurate and precise data for synoptic observations. Another point with suspected inhomogeneity appears in 1985 (dating back in 1981), but only in RDI series for wind direction.

In order to confirm detected break points, it is highly recommended to run the ReDistribution Test with different data window size. Confirmed break point should appear at more than one data window sizes with variations of one to two weeks with delay of up to two months. For example, such multiple runs of RDI values confirm detected break point in homogeneity dating in 1978 (Fig. 4). All RDI passes indicate the same point of inhomogeneity. Also, suspected break point for wind direction dating in 1981 is now confirmed. Series in Fig. 4 are shifted back to the 1-year moving values in order to illustrate this multiple pass indication.

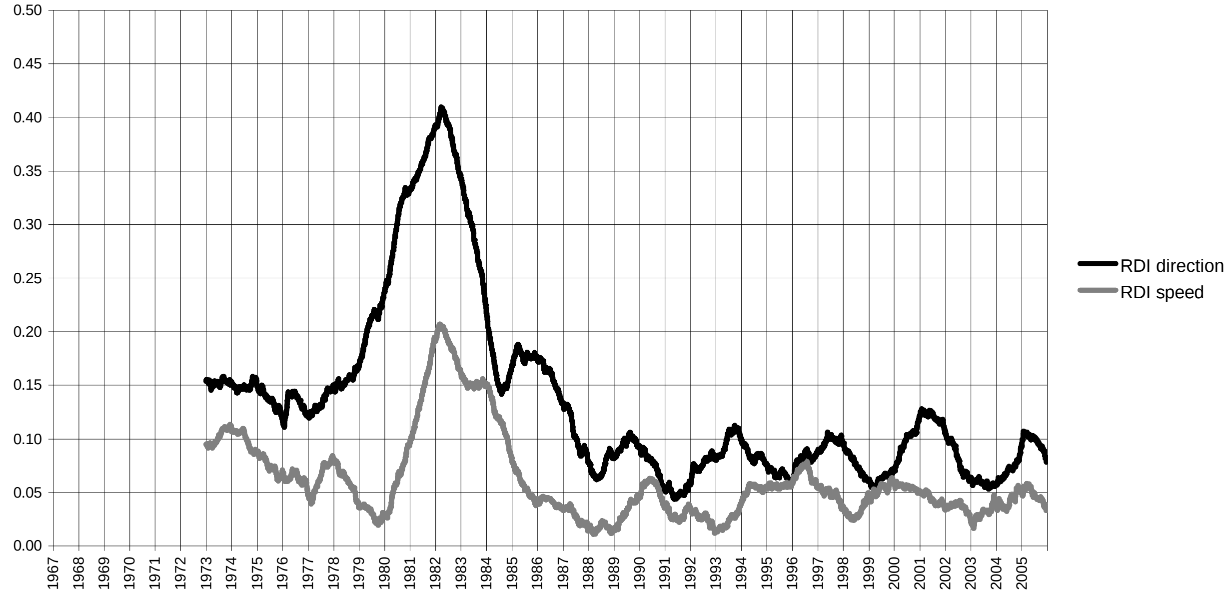


Fig. 3. ReDistribution Index (RDI) series for wind direction and speed, moving 4-year window period, Novi Sad - Rimski Šančevi, 1967-2005

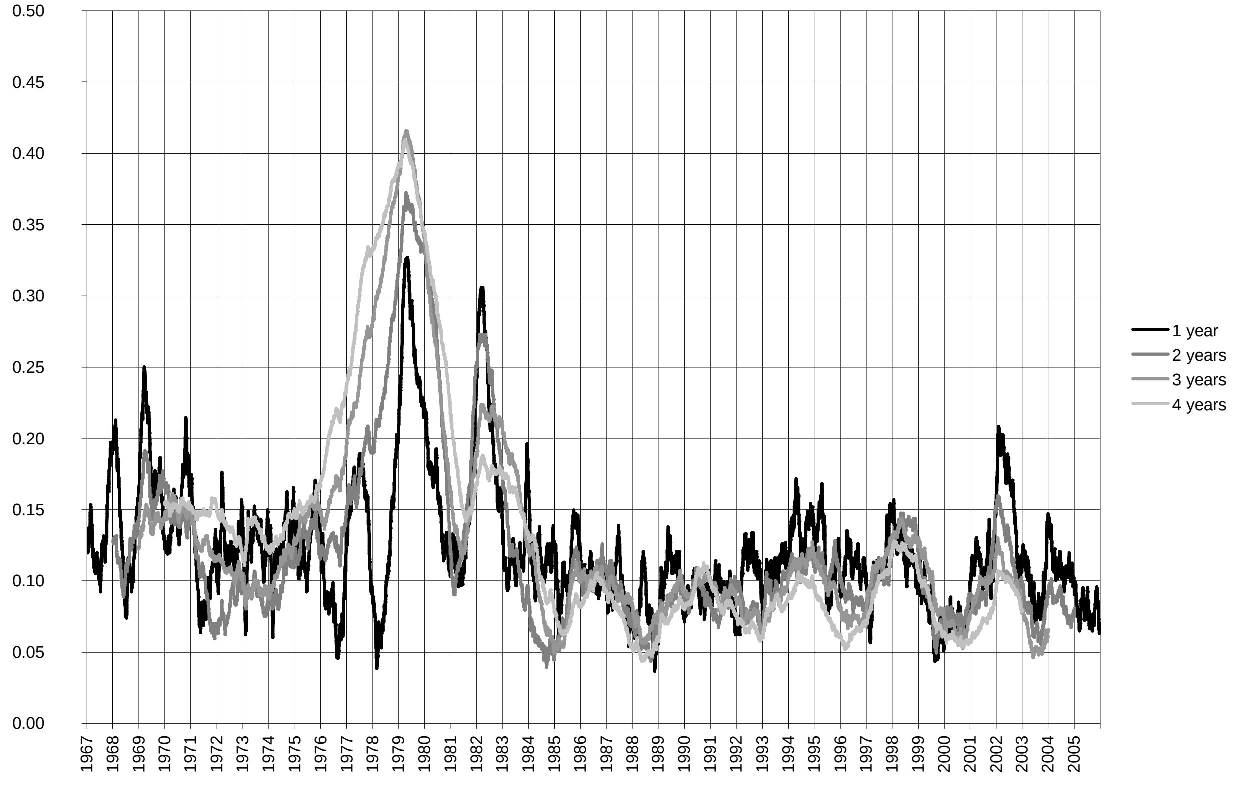


Fig. 4. Multiple passes of ReDistribution Index (RDI) series for wind direction and speed, moving 1, 2, 3 and 4-year window periods, Novi Sad - Rimski Šančevi, 1967-2005

Further data analysis might show the type of problem regarding wind data. Such analysis include displaying of wind direction frequencies ("wind roses"), or wind speed frequencies (by defined intervals - data classes), calculation of mean wind vector (or its components) for two consequential window periods. These types of information might be of great help in determing causes of inhomogeneity as well as in selecting ways for homogenisation of series.

4. TYPES OF INHOMOGENEITIES

According to occurences of ReDistribution Index (RDI) peaks in direction and / or speed series, there are three basic types of inhomogeneities of wind data.

Type 1. RDI values of **both** series have peaks at the same time. This practically means that both direction and speed data have simultaneous homogeneity breaks.

Possible causes for this type of inhomogeneity are:

Change of instrument. Such change might consider a change of instrument type (as in the given example), where major instrument properties (precision, sensitivity etc.) might be completely different. This also might be an indication of replacement of an instrument. Such information is usually included with the metadata.

Change of location / surroundings. This includes both relocation of instrument (position and / or height) and changes of surrounding objects (i.e. tree growth or cut, buildings or other objects arised or lowered down). Metadata might be incomplete regarding this problem, especially in changes of surroundings.

Change of observer. This might occur in observations that are not taken from the wind recording devices. Objectivity of the observer might be questioned in such cases, because at least one observer (before or after the detected break point) had no correct measurements. Still, it is almost impossible to prove the correctness of the data.

Homogenisation of the series is a set of various mathematical corrections. Since the source values from the instruments might be derived from almost entirely new conditions (instrument technical properties, position, different obstacles in surroundings), estimated correction value / function (if any) should incorporate a lot of calculations, where uncertainties might overwhelm the correction value itself. Therefore, homogenisation of such series is **not recommended**.

Type 2. Only RDI values of **wind direction** series have peak (not joined by the other RDI peak).

Possible causes for this type of inhomogeneity are:

Change of instrument orientation. This is usually case when an instrument orientation is corrected to geographical coordinates (or if it is, in seldom cases, miscorrected from geographical coordinates). Misorientation of instrument might happen at incorrect installation of instruments (i.e. when magnetic north is used as a reference direction), but it might also happen when the instrument is not properly maintained.

Change of measurement precision. Various instrument types have different number of distinguishable wind directions (there might be 8, 16, 32 or 36 directions plus calms). Such changes are seen as distortions of "wind roses". The "wind roses" with fewer numbers of directions look more "starry" when introducing those with more directions (Petrović, 2000). Thus, "starry" roses are not always showing the true distribution of wind direction frequencies.

Homogenisation of the series is possible only in some selected cases. For example, case of instrument orientation ("rotation" of the wind rose) should have a correction of azimouth value for a certain angle (Fig. 5 and 6). Detection of angle might be more accurate when the resulting wind vector is calculated before and after the break point. On the other hand, such accuracy is not quite practical for application, since there is a limitation of precision in detecting wind direction. Distortion of "wind roses" might be homogenised, but only if the result is the wind rose with fewer directions. In such case, precise wind data information is lost, and there is no point in homogenisation.

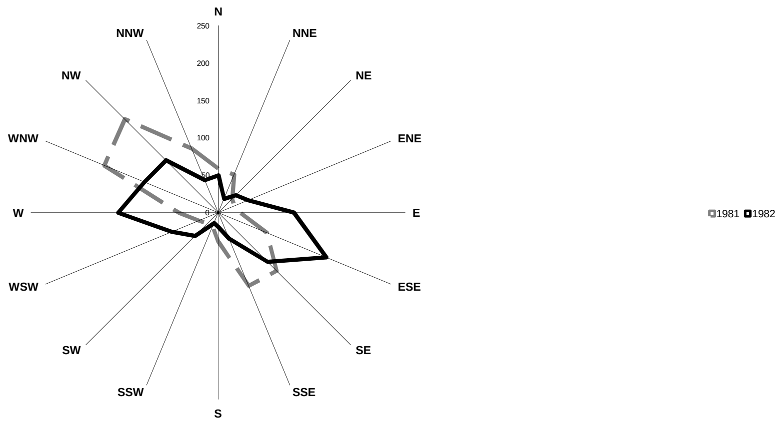


Fig. 5. "Wind roses" between two consequential moving years (1981-1982), original values, Novi Sad - Rimski Šančevi

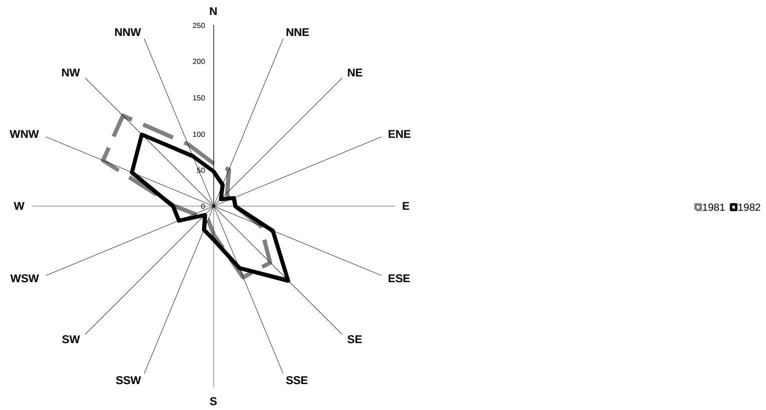


Fig. 6. "Wind rose" between two consequential moving years (1981-1982), corrected values, Novi Sad - Rimski Šančevi

Type 3. Only RDI values of **wind speed** series have peak (not joined by the other RDI peak). This practically means that the wind speed instrument is changed, while wind vane was replaced correctly or even remained intact.

Possible causes for this type of inhomogeneity are:

Change of instrument calibration. Like any other instrument, wind speed sensor must be calibrated for the correct values. In time, due to changes in friction of mechanical parts of the instrument, the values are smaller than true values. Therefore, the instrument must be calibrated again or replaced. Changes of instrumentation are usually featured in the metadata, but the genesis of the problem is almost never actually recorded.

Change of instrument sensitivity. Due to the same reasons, instrument sensitivity to low wind speed might change, so the threshold value for initialisation of wind speed instrument might increase. However, recorded high wind speed values might remain unchanged. As in previous case, some information might be available in the metadata, but not the genesis of the problem with instrument.

Homogenisation of the series is possible only in some selected cases. For example, in case of instrument calibration, homogenisation should be performed the same way like any other scalar value. On the other side, changes of instrument sensitivity usually have an unknown number of cases with low wind speeds replaced with zero values (and reversly). As a result, such case brings an unknown correction value and homogenisation is not recommended.

5. USE OF METADATA

True detection of inhomogeneities must include the use of metadata. Detected break point should be verified by a search through metadata for the true cause of inhomogeneity (Aguilar et al., 2003).

However, metadata might be incomplete in most cases. Old metadata often have an incomplete information about instruments, especially their technical features. Detailed descriptions of locations, surroundings, observers and sometimes even station locations are questionable because of the possible losses of metadata (i.e. damaged, destroyed or "lost" in some other country). In many cases there was no practice of recording changes of surroundings of the observation site.

Even with presumption of having the complete technical coverage in metadata, it is almost impossible to find all detailed descriptions of the station surroundings. While some of such descriptions might be discovered with buildings, tree growth or cut is practically impossible for detection. Moreover, it is still difficult to estimate direct influences of surroundings on the instrumentation.

Despite the development of homogenisation techniques, there is still a significant number of detected break points that are not verified in metadata, hence the cause of the detected inhomogeneities remains unknown (Auer et al., 2003, Müller-Westermeier, 2003). Some metadata might be partly recovered by using other techniques that attribute data quality, such as the Real Precision Method (Petrović, 1998). These techniques might discover a "hidden" information on the observers and reliability of observations. Therefore, discovering new facts is quite useful for completion of metadata.

6. FURTHER ACTIONS

Since some causes of inhomogeneities have influence on more than one weather element, detection of inhomogeneities should be performed on as more elements as possible. Matching break points clearly indicate major changes at an observation site (i.e. station relocations) and determine their influences on observation records.

The ReDistribution Method offers many possibilities for detection of various changes in observations. Any series that might have an empirical distribution of its values might be processed in order to detect new inhomogeneities (like any other homogenisation test) or to obtain more information for completion of metadata. This also might include elements whose "homogeneity" makes no scientific sense (i.e. visibility, start / end time of observing weather phenomenae). Some preliminary results are encouraging for evaluation this method with such series, but only as a tool for detection and hence estimation of missing metadata. Discoveries and assumptions of new facts about a weather station are always of great help in searching for new inhomogeneities.

REFERENCES

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., Wieringa, J., 2003: Guidelines on climate data and homogenization, WMO, WCDMP-No. 53

Auer, I., Böhm, R., Scheifinger, H., Ungersböck, M., Orik, A., Jurković, A., 2004: "Metadata and their role in homogenising", Fourth Seminar For Homogenization And Quality Control In Climatological Databases (Budapest, Hungary, 6-10. October 2003); WMO, WCDMP-No. 56, pp. 17-23

Müller-Westermeier, G. 2004: "Statistical analysis of results of homogeneity testing and homogenisation of long climatological time series in Germany", Fourth Seminar For Homogenization And Quality Control In Climatological Databases (Budapest, Hungary, 6-10. October 2003); WMO, WCDMP-No. 56, pp. 25-38

Petrović, P. 1999: "Measurement precision as a cause of inhomogeneity in weather data time series", Proceedings From The Second Seminar On Homogenization Od Surface Climatological Data, (Budapest, Hungary, 9-13 November 1998), WNMO, WCDMO, pp. 161-169

Petrović, P. 2001: "Some results of exploration and development of the Real Precision Method", From The Second Seminar On Homogenization Od Surface Climatological Data, Budapest

Petrović, P. 2004: "Detecting of inhomogeneities in time series using Real Precision Method, Fourth Seminar For Homogenization And Quality Control In Climatological Databases (Budapest, Hungary, 6-10. October 2003); WMO, WCDMP-No. 56, pp. 79-88

URBAN EFFECTS ON THE TEMPERATURE TIME SERIES OF PRAGUE

Vít Květoň, Michal Žák
Czech Hydrometeorological Institute, Prague
vit.kveton@chmi.cz, michal.zak@chmi.cz

1. INTRODUCTION

An urban heat island is a metropolitan area which is significantly warmer than its surroundings. As population centres grow in size from village to town to city, they tend to have a corresponding increase in average temperature, which is more often welcome in winter months than in summertime. This phenomenon is called the "urban heat island effect." It is caused by larger absorption of shortwave solar radiation due to greater absorbing capacity of buildings, roads etc. This radiation is then stored in these materials and later, mainly during night, it is irradiated in the form of long-wave radiation and therefore it is warmer in the inner cities than in their surroundings. If we study an urban heat island, we usually try to determine the intensity of urban heat island. Intensity of urban heat island appeares in the difference in temperature of the city and outside, of the rural areas.

2. DATA USED FOR ANALYSIS

Data for time period of 1961-2005 have been used for analysis. Prague temperature time series at following stations were used: Klementinum, Karlov, Ruzyně, Kbely, Libuš and Uhříněves. For assessing of urban heat island intensity, following rural stations have been used: Doksany and Ondřejov. The position of these all used stations can be seen on the Fig. 1.

For the analysis, we have used following temperature characteristics: annual average of daily maximum, annual average of daily minimum and average air temperature. Further, annual number of the so called „characteristic“ days have been used. As “characteristic” day we mean day with temperature above given threshold. We have used the following characteristic days:

- Tropical days that are days with maximum air temperature 30°C and more;
- Summer days that are days with maximum air temperature 25°C and more;
- Frost day that are days with minimum air temperature below 0°C;
- Ice day that are days with maximum air temperature below 0°C.

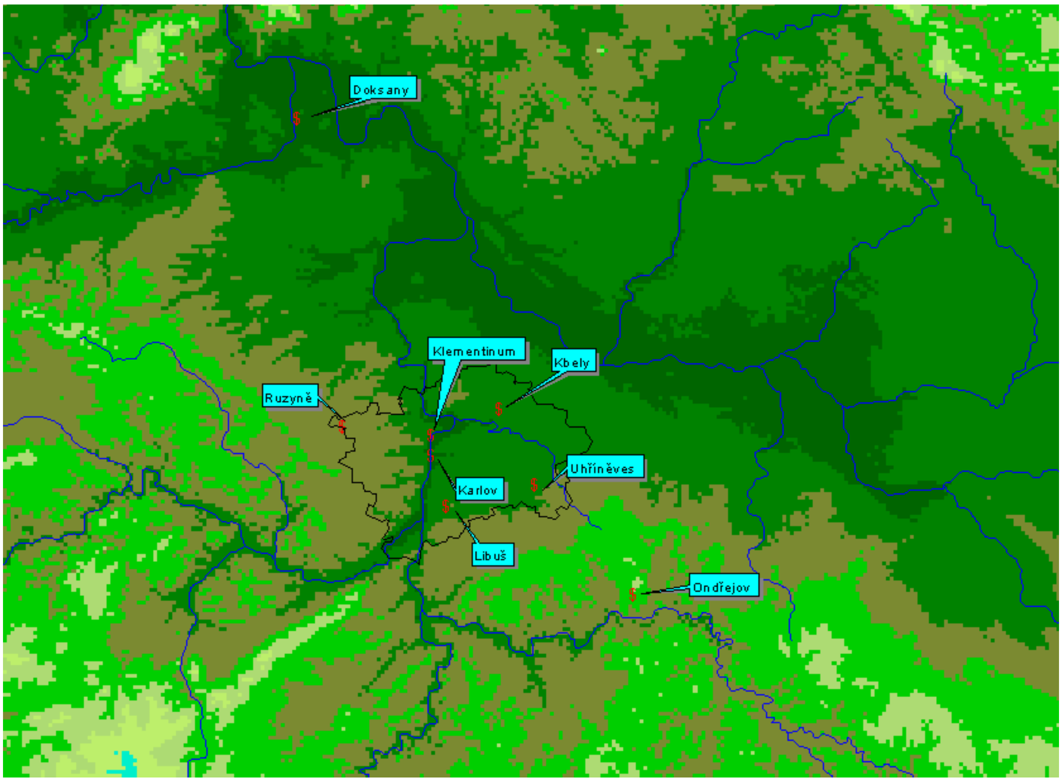


Fig. 1: Map of used station, black line is border of Prague

3. HOMOGENIZATION OF DATA

Of course, all data we have used had to be homogenized. For this purpose we have used MASH method (see e.g. [1]) to detect main break points and/or shifts and their approximate size. Then we have used auxiliary graphs of cumulative sums of relative differences of annual temperatures from average 1961-2005 according to followed formula:

$$\sum((ai\text{-}a)/a - (bi\text{-}b)/b)), \qquad where\ i = 1, \ldots, 45,$$

ai, bi are annual temperatures in i-th year, a, b are average temperatures for the period 1961-2005 on the tested station and the regional station REF respectively. A letter a is related to the tested station, a letter b to the station REF of given station group. An example of such graph for average air temperature can be seen on Fig. 2 for original data and on Fig. 3 after homogenization. As we can see, these sums are 5 to 10 times smaller after homogenization.

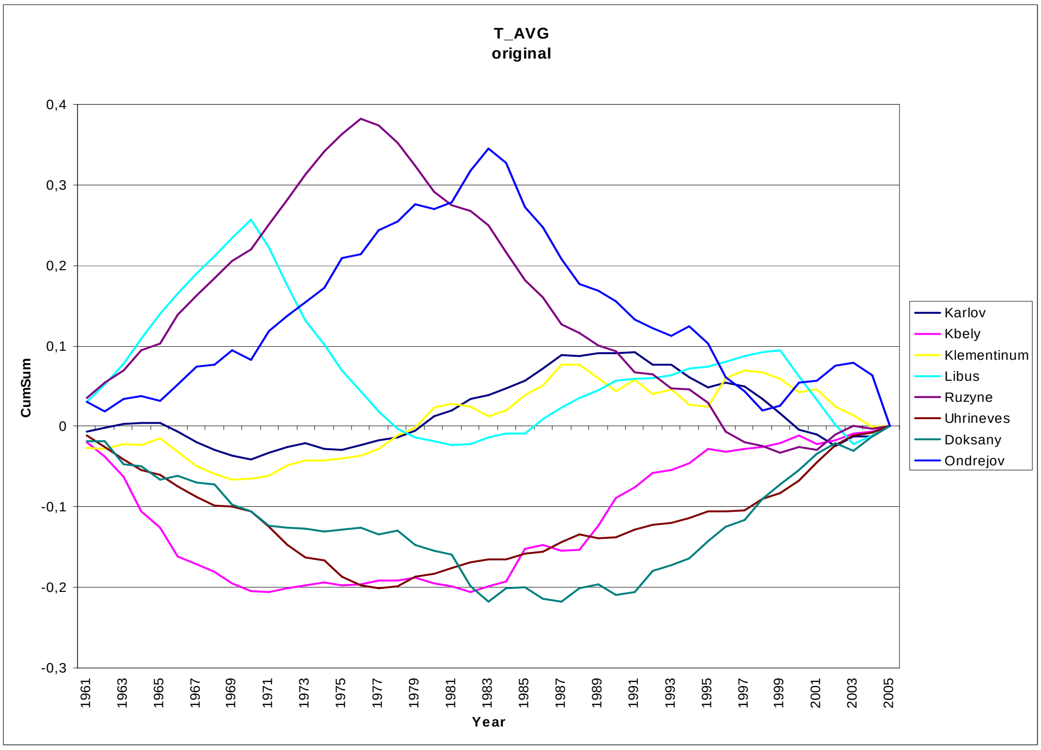


Fig. 2: Auxiliary graph of cumulative sums of relative differences of annual air temperatures – before homogenization

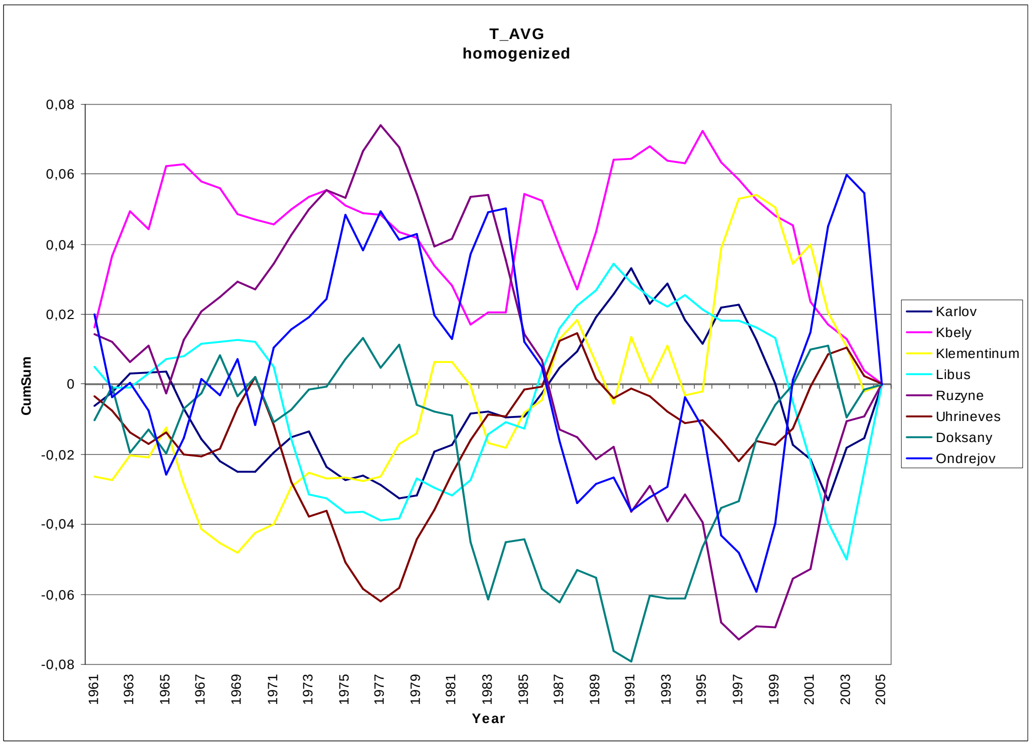


Fig. 3: Auxiliary graph of cumulative sums of relative differences of annual air temperatures – after homogenization

4. RESULTS

For exploring the intensity of urban heat island, we used graphs showing variations of differences in temperatures for years 1961-2005. The situation for average temperature can be seen on Figs. 4 and 5 for two examples, first for differences between Prague station Klementinum and rural station Doksany (that are almost in the same altitude), and the second one shows the differences between Prague periphery station Ruzyne and rural station Ondrejov (where Ondrejov lies some 100 m higher then Ruzyne). From both graphs it can be seen that the influence of urban heat island has intensified during last 45 years, this intensification is smaller when compared Klementinum in the centre of Prague with rural station Doksany (intensification of about 0.06°C/10 years) then in the periphery (Fig. 5, intensification of about 0.3°C/10 years).

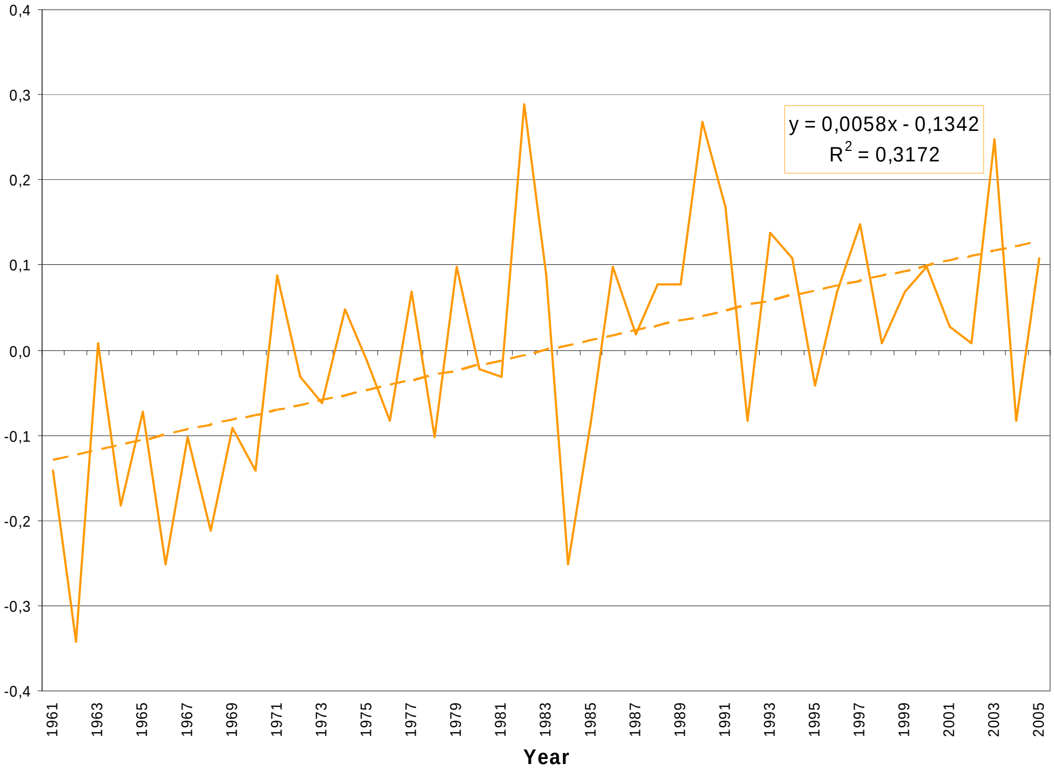


Fig.4: Variation of differences in average temperature for stations Klementinum and Doksany

Focusing now on the minimum temperature, situation for the same pairs of stations can be seen on Figs. 6 and 7. As for the historical centre of Prague, there is relatively large intensification of urban heat island (about 0.18°C/10 years), but no changes can be seen for Ruzyne-Ondrejov graph (Fig. 7). As for maximum temperatures, the situations are shown on Figs. 8 and 9, the influence of the urban heat island intensity is only very small for the historical centre and no evidence of it can be detected on periphery of the city.

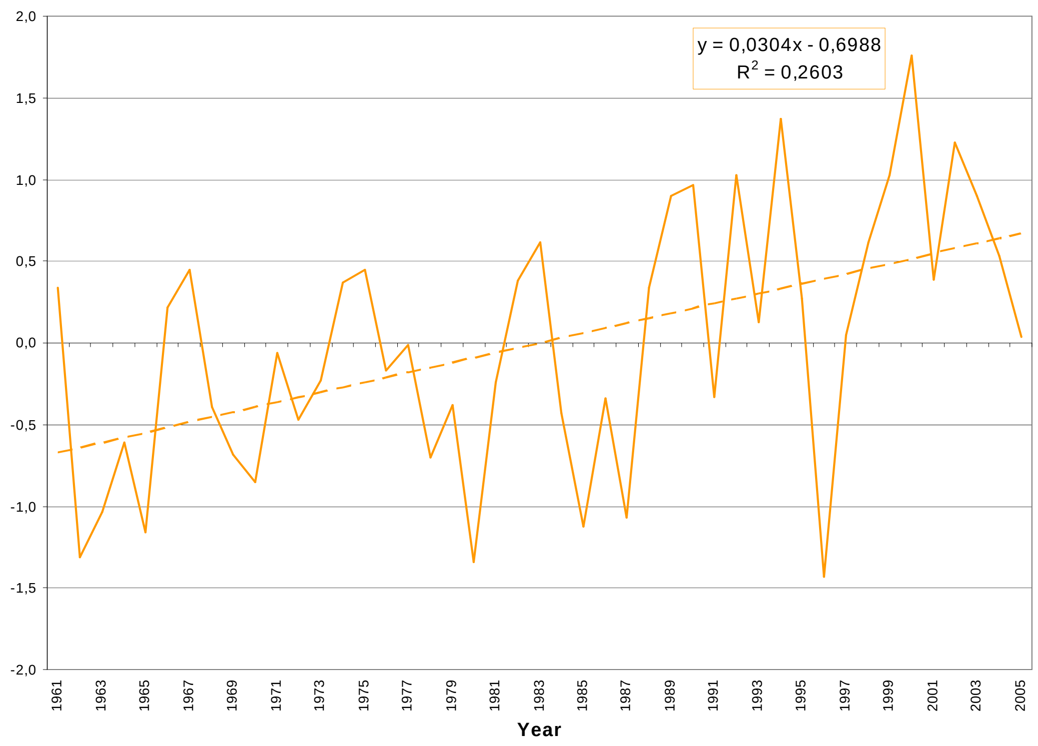


Fig. 5: Variation of differences in average temperature for stations Ruzyne and Ondrejov

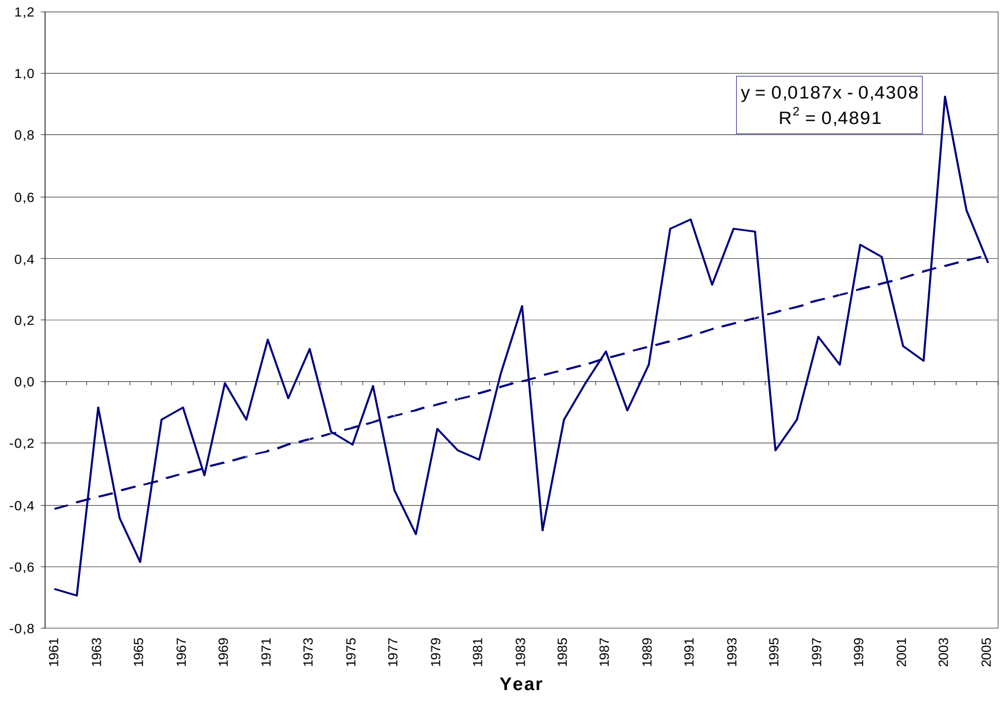


Fig. 6: Variation of differences in minimum temperature for stations Klementinum and Doksany

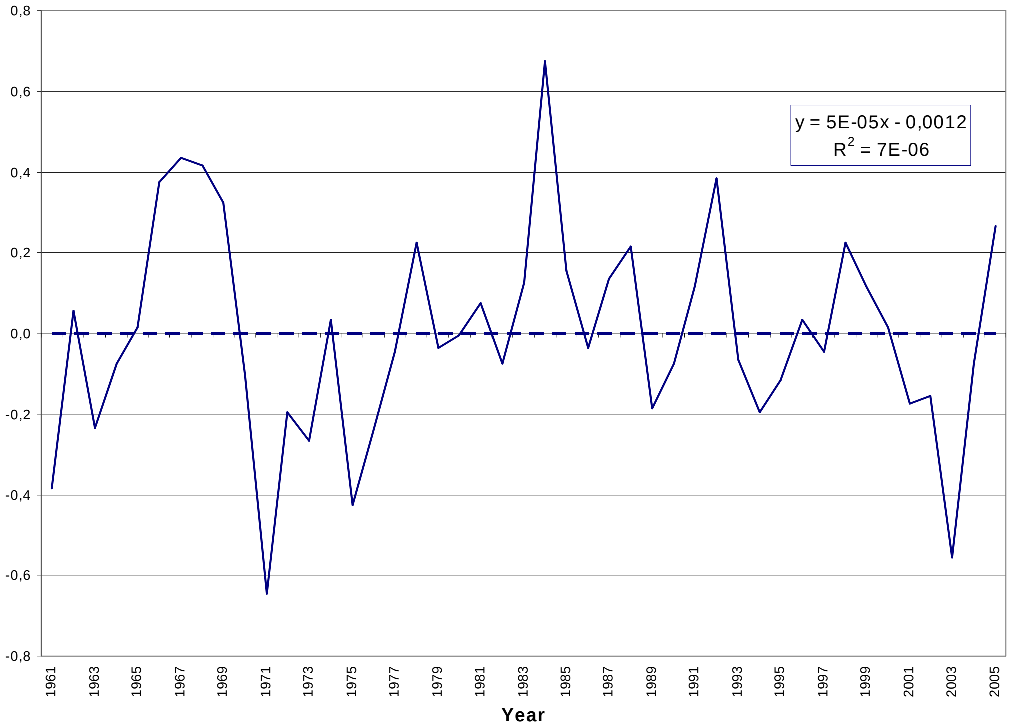


Fig. 7: Variation of differences in minimum temperature for stations Ruzyne and Ondrejov

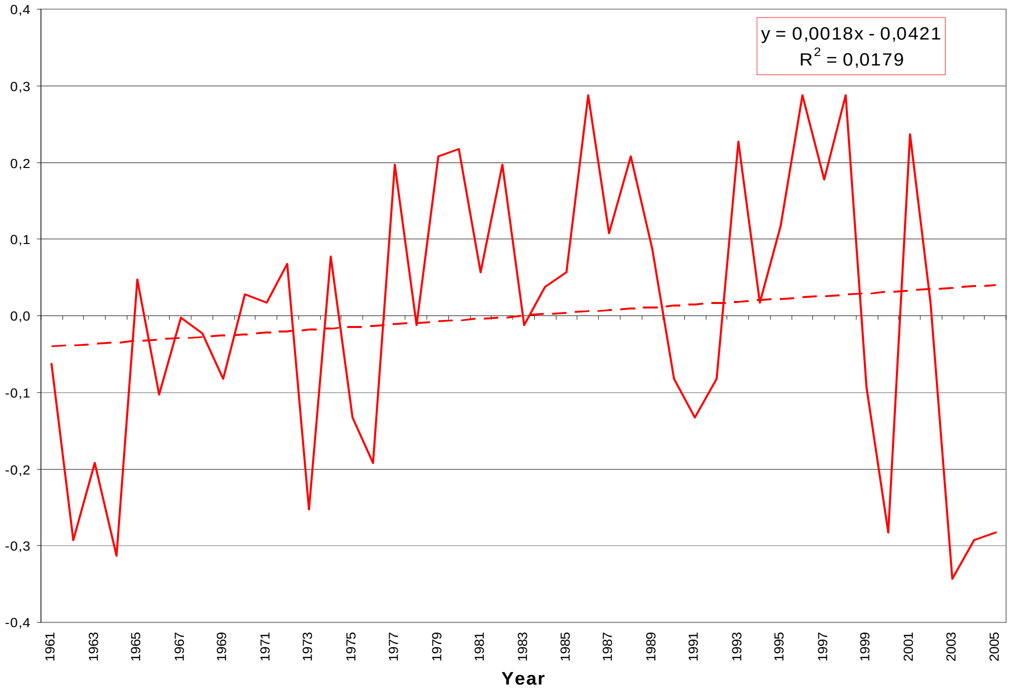


Fig. 8: Variation of differences in maximum temperature for stations Klementinum and Doksany

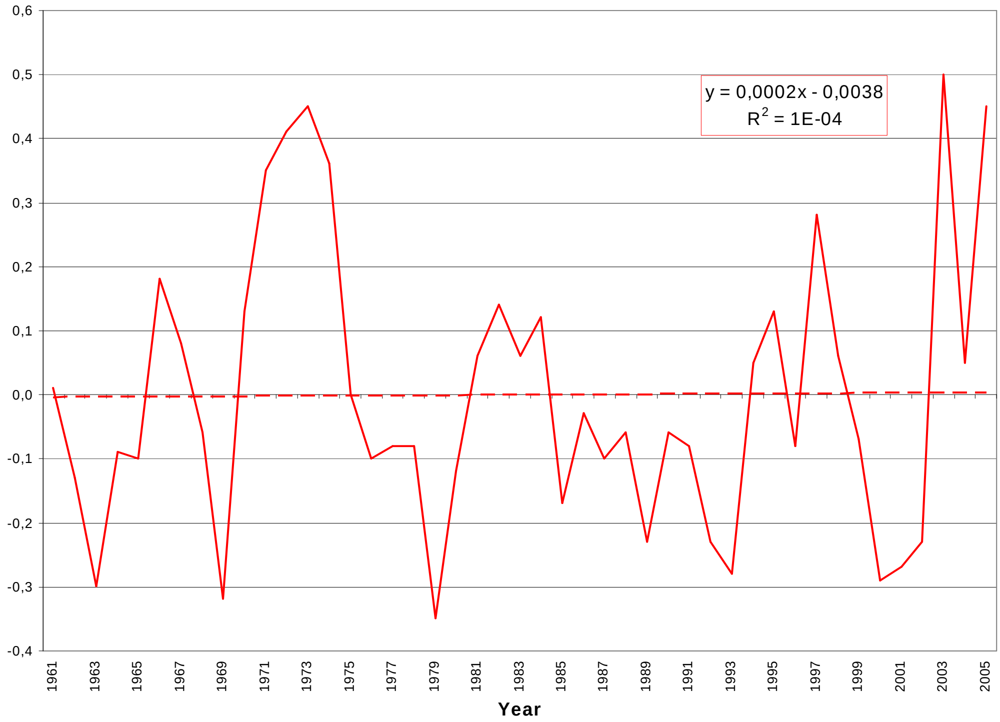


Fig. 9: Variation of differences in maximum temperature for stations Ruzyne and Ondrejov

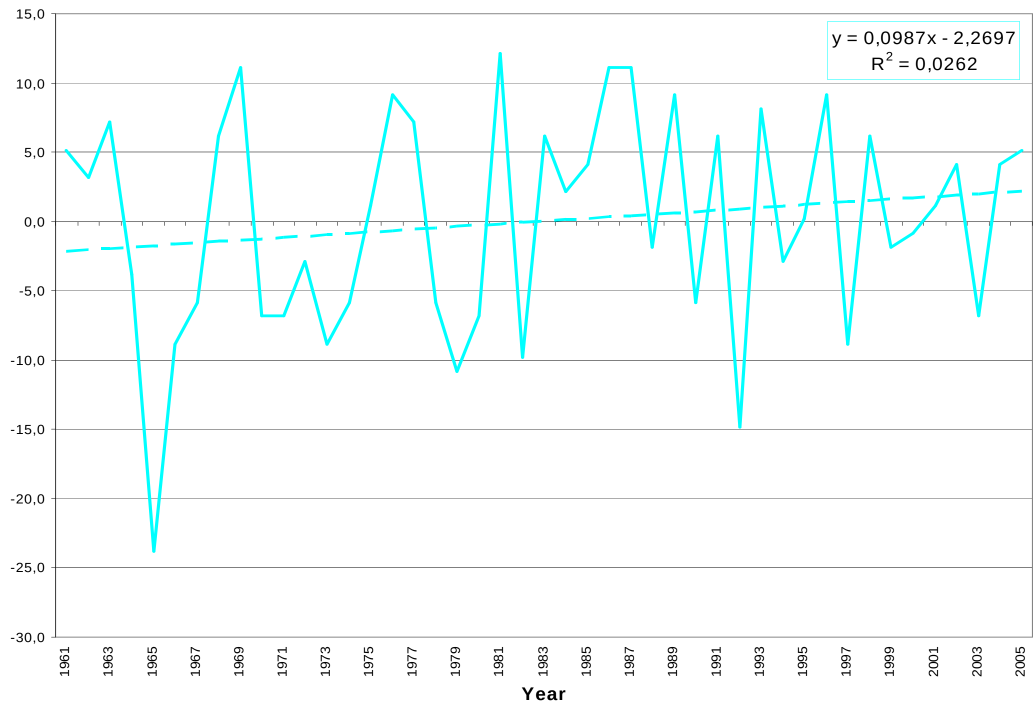


Fig. 10: Variation of differences in number of frost days for stations Klementinum and Doksany

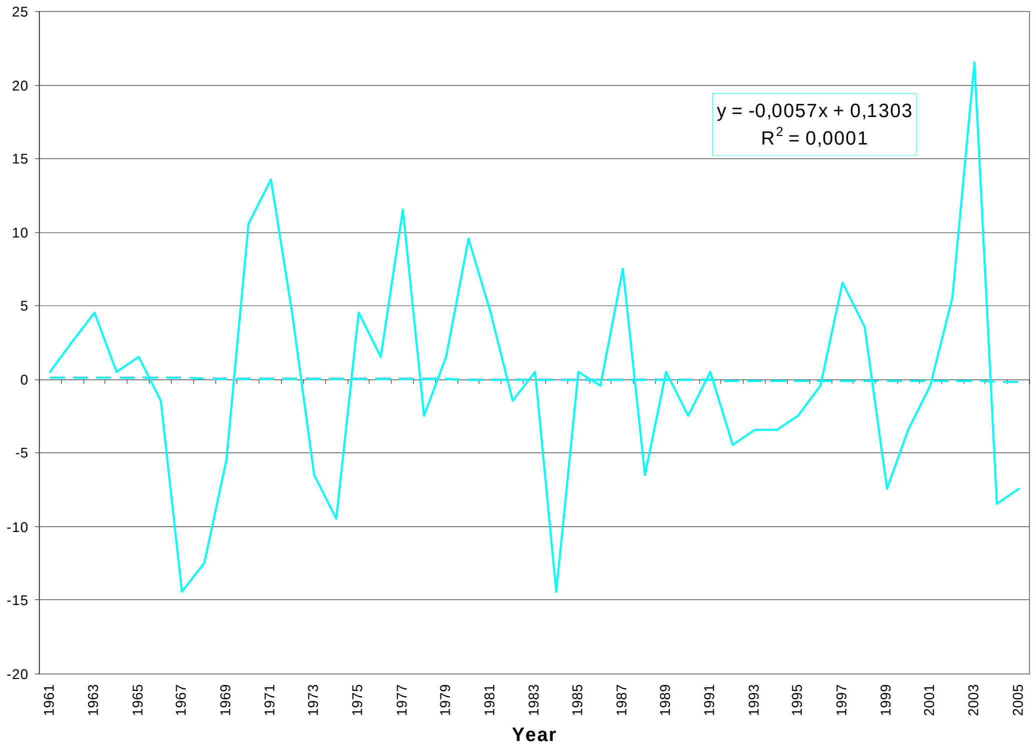


Fig. 11: Variation of differences in number of frost days for stations Ruzyne and Ondrejov

Finally, graphs on Figs. 10 and 11 present variations of differences in number of frost days. We can see that also here the differences are significant only for the city centre. But now the situation is somewhat different. The number of frost days generally decreased during the last 40 years, so the increasing differences between Klementinum and Doksany mean that the number of frost days in the inner city decreases more slowly then outside of the city.

The above mentioned results can be well demonstrated and summarized on the graphs showing the differences of trends of temperatures between Klementinum and other stations. The columns in these graphs are ordered according to increasing distance from the city centre (Klementinum - 0 km). These graphs are presented on Fig. 12 for temperatures (average, minimum and maximum) and for number of characteristic days (tropical, summer, frost and ice). We would expect that the influence of the urban heat island decreases with increasing distance from the city centre. It means that we would also expect to increase these differences in trends with distance. As we can see for temperatures it does so, but with some exceptions. These exceptions are caused by generally inhomogeneity of the urban heat island in various directions. Finally, for characteristic days (Fig. 13), the situation is not as clear or easy as in the case of temperature itself. Generally the number of characteristic days decreases and this decrease is growing considerable with the increasing distance from the city centre.

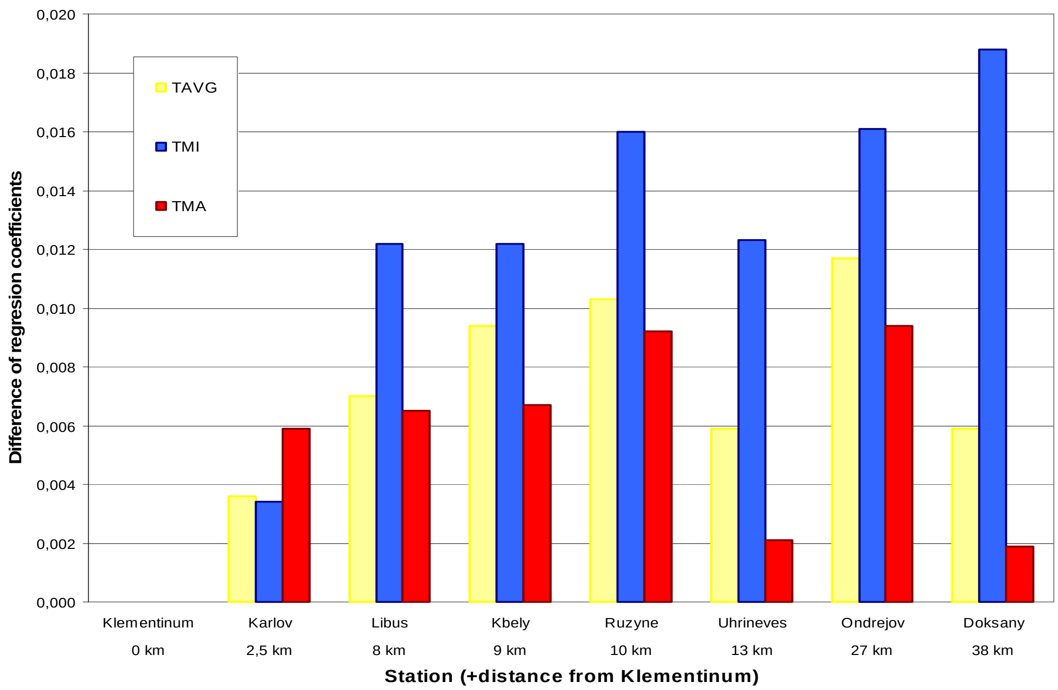


Fig. 12: Trends of temperature, differences of Klementinum and other stations, period 1961-2005

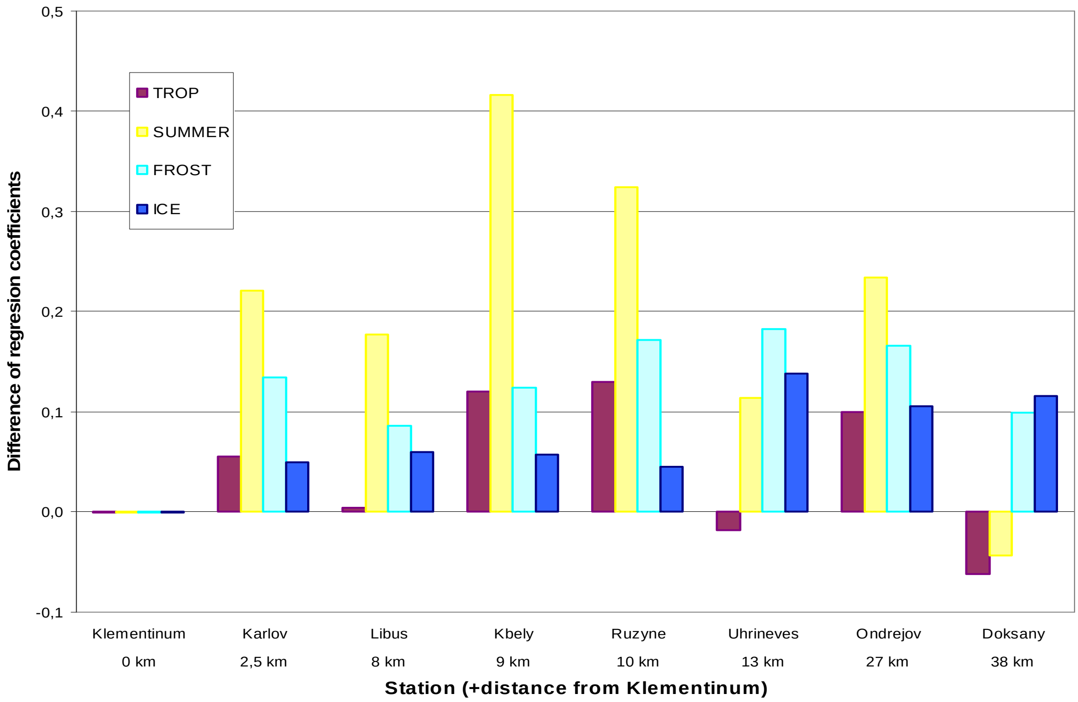


Fig. 13: Trend of temperature, differences of Klementinum and other stations, period 1961-2005

5. SUMMARY OF RESULTS FOR 1961-2005

We have found relatively large intensification for minimum temperatures (0.13-0.15°C/10 years) caused by urban heat island, as well as for average air temperatures (0.06°C/10 years), but only small intensification in maximum temperatures (0.02-0.03°C/10 years). The changes in number of characteristic days are not very helpful in the sense of urban heat island intensity changes. Number of summer days in rural station Doksany is similar to the station Prague Klementinum. This could be due to special location of Doksany station that expresses by above-average temperature under warm weather conditions. We could detect increase of tropical days (1 day/10 years). Number of ice and frost days decreases in the periphery and outside of the city faster than in the inner city. This could be caused by the influence of building heating, better ventilation, etc.

Finally, let's mention some more general remarks about Prague heat island by other authors. Huth and Beranova [2] studied Prague heat island under different synoptic conditions. They found that maximum urban heat island occurs in summer, whereas minimum occurs in winter. The highest frequency of urban heat island occurs under windy weather conditions with north-northeastern components. The highest intensity of Prague heat island (about 2.58°C) can be observed under anticyclonic weather conditions and when north-northeastern winds are prevailing. Besides, they have found that long-term trend of increasing intensity of heat island is about 1.2°C/100 years, when the highest trend can be detected under north-northeastern and south-southwestern wind weather conditions (2.2°C) and no trend can be detected under east-southeastern and west-northwestend wind weather conditions.

Brazdil and Budikova [3] have studied period of 1921-1995 and they have found positive and statistically significant trends documenting additional warming of the Klementinum station with the most conspicuous and significant warming in winter and spring (0.06°C/10 years). It is similar to what we have found for average air temperature. This warming is well correlated with growth of the population (from 150 thousands in 1850 to 1.2 millions in 2000), consumption of energy and expansion of the built-up area in Prague

6. CONCLUSIONS

We have found well pronounced urban effects on the temperature time series in Prague, mainly in the inner city. But we have also found some problems that could be caused by influence of local conditions on measurements (representativeness of location, e.g. thermometer in Klementinum is situated 6 m above ground, closed space of the courtyard). Another question is the eventual insufficiency of homogeneity of temperature time series, mainly in number of characteristic days. Therefore more detailed study should be made to solve these problems and answer these questions.

REFERENCES

[1] Szentimrey, T., 2002: "MASHv2.01", Guide for Software Package, Hungarian Meteorological Service, Budapest, Hungary, p. 42.

[2] Beranova, R., Huth, R., 2003: Prague heat island under different synoptic conditions (in Czech only), Meteorologicke zpravy, 56.

[3] Brazdil, R., Budikova, M., 1999: An urban bias in air temperature fluctuations at the Klementinum, Prague, Czech Republic. Atm. Environ., 33, 4211-4217.

HOMOGENIZATION OF THE SPANISH DAILY TEMPERATURE SERIES, A STEP FORWARD

Aguilar, E., Brunet, M., Saladié, O., Sigró, J.
Climate Change Research Group. Geography Unit
Universitat Rovira i Virgili de Tarragona,
Plaza Imperial Tarraco, 1, 43005, Tarragona, Spain
enric.aguilar@urv.net

ABSTRACT

During the last decade, the Climate Change Research Group (CCRG) has devoted its efforts to the compilation of different temperature datasets for Spain. Following the trends in global research, our focus has changed from monthly to daily data. The recent release of the Spanish Daily Adjusted Temperature Series (*SDATS*, Brunet *et al*., in press) constitutes a great advance for the analysis of climate change in the Iberian Peninsula during the instrumental era, including 22 long-term stations, suitable for climate change and variability analysis.

Our homogenization methodology combines several techniques for an optimum adjustment of monthly and daily temperatures. For the adjustment of monthly data, our scheme relies on the application of direct and relative homogenization techniques. Direct homogenization is applied to the late 19th century and early 20th century data, where almost all the network was impacted by the change from open stands (basically Montsouris) to the Stevenson Screen. In these conditions, relative homogenization proves to be inefficient, so two replicas of the ancient stands were constructed and placed in Murcia & La Coruña, next to the official Stevenson Screens of the Instituto Nacional de Meteorologia. The paired observations recorded during almost three years, were used to obtain correction factors for the so-called "screen-bias", previously to the application of the Standard Normal Homogeneity Test Daily data is adjusted by linear interpolation of the monthly factors.

1. INTRODUCTION

One of the principal needs for climate analysis is the availability of high quality and homogeneous data series. The observed fraction of the climate data does not only allow us to describe the last couple of centuries of the globe's climate, but also plays a crucial role in the calibration of proxy records and models. For these reasons the reconstruction, quality control and homogenization of climate time series constitutes an important effort to be overtaken by the climate community.

Many methods have been successfully applied by different groups of scientists to homogenize annual to monthly values. Translating monthly homogenization factors to daily adjustments is a difficult task (Aguilar *et al*., 2003), although great improvements have been made recently (for example, Della Marta *et al*., 2006). The CCRG current approach consists of the interpolation of the monthly factors into daily values using the effective approach described by Vincent *et al*. (2002), which efficiently accounts for the annual cycle between months and allows obtaining reasonably homogeneous series for most proposes.

The rest of this paper is structured in 4 additional sections. Section 2, describes the Spanish Daily Temperature Series (Brunet *et al*., in press) and the quality control procedures applied; section 3 deals with homogenization procedures; section 4 discusses our procedures and makes conclusions; section 5 lists the quoted references.

2. DATA AND QUALITY CONTROL

The Spanish Temperature Series is composed by the 22 most reliable long-term stations (see table 1) of daily maximum and minimum temperature. The longer series start back in 1850, although a few of them do not have data until the first decade of the 20th century. Data is kept up to date with new incoming values and the stations' distribution permits a good coverage for the analysis of multidecadal temperature variability and change in Spain (see figure 1).
Although most of the data was kindly facilitated by the Spanish Instituto Nacional de Meteorología, the CCRG did an important effort on data archaeology and digitization, recovering important sections of the series, that were lost for the climate community until that time (see Brunet *et al*., *in press*). The total amount of data now available ads up to around 2 million maximum and minimum temperature values.
All the ingested data, with independence of its procedence, was quality controlled with the application of the following tests:

- ⇒ Comparison of original source monthly mean and dataset monthly mean
- ⇒ Values out of [-50,50]ºC interval
- ⇒ Tmax < Tmin
- ⇒ Number of days per year, month
- ⇒ Duplicates
- ⇒ Excedence of 4SD
- ⇒ Interdiurnal differences > 25ºC
- ⇒ 4 or more consecutive identical values
- ⇒ Excedence of 4SD of difference between candidate and reference series
- ⇒ Visual comparisons among neighbours

Only a small fraction of data where flagged (0.58%), although they resulted in a considerable number of individual checks (roughly 12,000). After consulting the original sources, some 8,000 values were recovered and corrected, the rest remaining unrecoverable and lost to the series.

Table 1: Spanish Daily Temperature Series, SDATS. List of Stations.

Location	Longitude (º)	Latitude (º)	Elevation (m)	Period
ALBACETE	01º51'47''-W	38º57'08''-N	698.56	1893–2003
ALICANTE	00º29'40''-W	38º22'00''-N	81.5	1893–2003
BADAJOZ	06º49'45''-W	38º53'00''-N	185	1864–2003
BARCELONA	02º10'36''-E	41º25'05''-N	420.1	1885–2003
BURGOS	03º36'57''-W	42º21'22''-N	881	1870–2003
CADIZ	06º12'37''-W	36º27'55''-N	30	1850–2003
CIUDAD REAL	03º55'11''-W	38º59'22''-N	627	1893–2003
GRANADA	03º37'52''-W	37º08'10''-N	685	1893–2003
HUELVA	06º54'35''-W	37º16'48''-N	19	1903–2003
HUESCA	00º19'35''-W	42º05'00''-N	541	1861–2003
LA CORUñA	08º25'10''-W	43º22'02''-N	67	1882–2003
MADRID	03º40'41''-W	40º24'40''-N	678.9	1853–2003
MALAGA	04º28'57''-W	36º39'57''-N	6.54	1893–2003
MURCIA	01º07'14''-W	37º58'59''-N	57	1863–2003
PAMPLONA	01º38'21''-W	42º46'06''-N	452	1880–2003
SALAMANCA	05º29'41''-W	40º56'50''-N	789.8	1893–2003
S. SEBASTIAN	02º02'22''-W	43º18'24''-N	251.6	1893–2003

SEVILLA	05º53’47’’-W	37º25’15’’-N	31	1893–2003
SORIA	02º29’01’’-W	41º46’29’’-N	1083	1893–2003
VALENCIA	00º22’52’’-W	39º28’48’’-N	11.4	1864–2003
VALLADOLID	04º44’35’’-W	41º38’40’’-N	691.4	1893–2003
ZARAGOZA	01º00’29’’-W	41º39’43’’-N	245	1887–2003

3. HOMOGENIZATION PROCEDURES

The CCRG’s homogenization methodology (see figure 1) relies on the combination of three different techniques. Direct homogenization is needed to solve a network-wide problem, the change from ancient screens (mostly Montsouris Screens) to the standard Stevenson Screen. This is indispensable to avoid the underestimation of trends in daily maximum temperatures, as open stands are highly impacted by direct solar radiation.

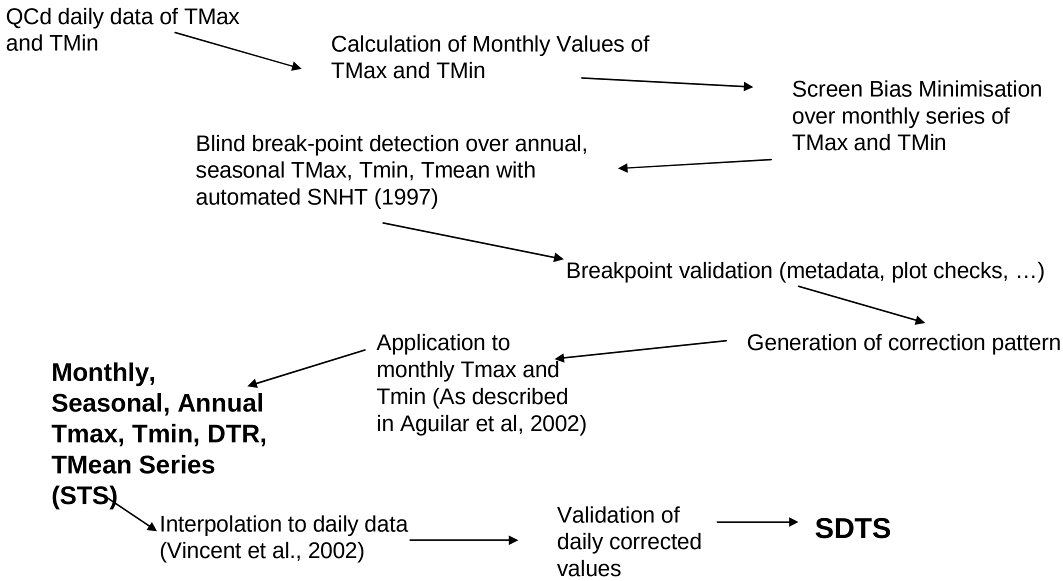


Figure 1: Homogenization procedures of the CCRG.

To remove the screen bias, the CCRG has built two Montsouris-replicas in the northwestern (La Coruña) and northeastern (Murcia) corners of the country, which were installed in 2003 nearby the Stevenson Screen in use of the official meteorological station (Brunet *et al.*, 2004). After having an almost 3 years long dataset, correction factors are derived by regressing Montsouris data to Stevenson data (see figure 2).

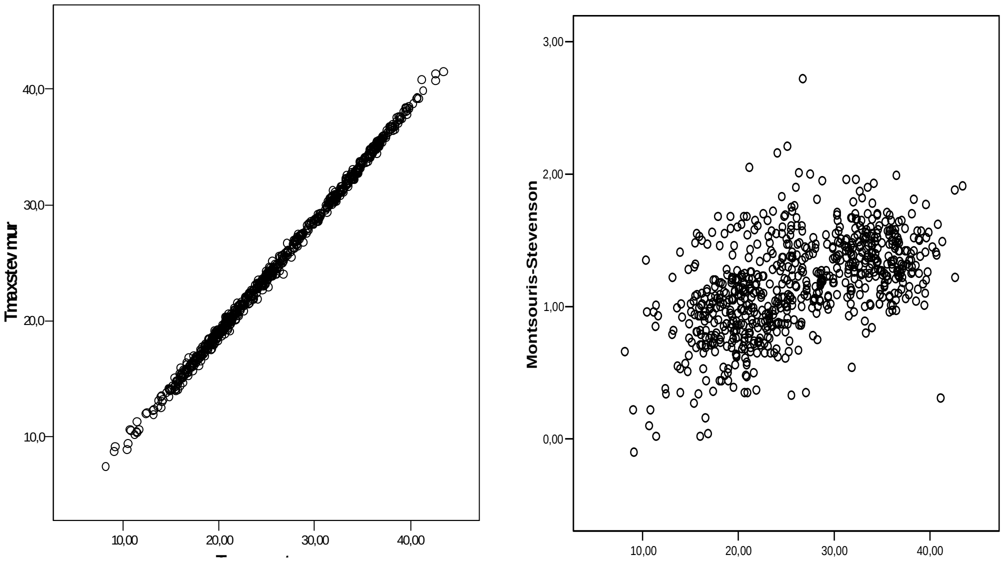


Figure 2: Empirical evaluation of the screen bias in Murcia.

After the screen bias minimization, the relative Standard Normal Homogeneity Test is applied to the series following the procedure described by Alexandersson and Moberg (1997), modified by Aguilar *et al.* (2002). For this propose, a blind run of the test is done for the whole dataset over the time series of annual averages of daily maximum temperature, daily minimum temperature, daily mean temperature and diurnal temperature range, obtaining a number of possible breakpoints. These potential inhomogeneities are checked against the available metadata and graphically analyzed by the inspection of the data and the z series (standardized difference of candidate-reference series). Breakpoints are then validated or rejected and a preliminary correction pattern is drawn. The *homogeneous* sections between accepted breakpoints are tested for artificial trends by evaluating the slope of the candidate-reference series. The definitive correction pattern is applied to the monthly averages of daily maximum and minimum temperatures. Diurnal temperature range and monthly means are derived from the homogeneous maximum and minimum series.

A third step is to translate the monthly factors into daily factors. This is achieved by the application of the interpolation described by Vincent *et al.* (2002), which avoids unnatural discontinuities at the end of the month and preserves the monthly averages. The monthly factor is assigned to the 15th day, and the factors for the rest of the month are lineally interpolated. As corrections for daily maximum and daily minimum temperatures are independent, a reduced number of daily maximum values are minor or equal to the corresponding daily minimum temperature. To correct this situation, they are forced to preserve the ration of change of the DTR from the original monthly series to the homogenized monthly series

4. DISCUSSION & CONCLUSSIONS

The CCRG's procedure for homogenization is a step forward to achieve reasonably homogeneous values of daily temperatures, suitable for most climate analysis. Although we acknowledge that our approach does not include any correction dependent on the particular weather of the day (i.e. different corrections should be applied to a sunny and a cloudy day) we do account for the annual cycle without compromising the monthly average, which is derived from a solid and well tested homogenization procedure of annual to monthly time series.

Despite of this, our approach allows the use of the SDATS in the calculation of extreme indices (see Brunet *et al.*, 2006; Moberg *et al.*, accepted) and its application to other datasets may improve data availability, and, of course, the analysis. Figure 3 shows an example of homogeneous indices calculated after the application of the CCRG methodology.

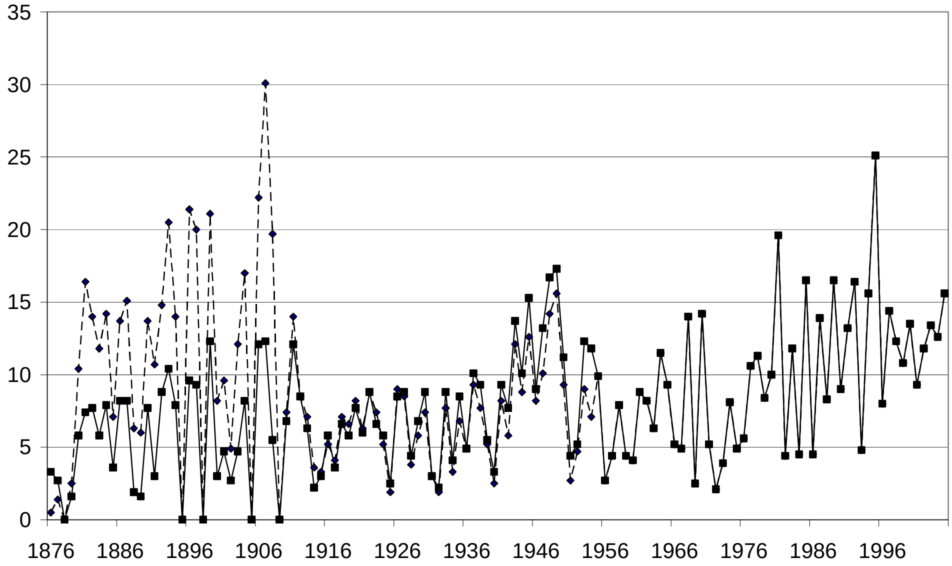


Figure 3: Annual Index TX90p (% of excedence of the 90th percentile of the 1961-1990 reference period).Dashed line: inhomogeneous index calculated over the original data.; solid line: homogeneous índex after the removal of the screen bias (1908) and the inhomogeneity produced by the relocation in 1954.

In conclusion, the SDATS represents a new tool for climate variability and change analysis in Spain, containing extended, quality controlled and homogenized daily records for the 22 stations having the longest series available.

REFERENCES

Aguilar E, Brunet M, Saladié O, Sigró J, López D. 2002. Hacia una aplicación óptima del standard normal homogeneity test para la homogeneización de series de temperatura. In La Informaci´on Climática Como Herramienta de Gestión Ambiental, VII Reunión Nacional de Climatología, Grupo de Climatología de la AGE, Cuadrat JM, Vicente SM, SAZ MA (eds). Universidad de Zaragoza: Zaragoza, 17–33.

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidelines on climate metadata and homogenization. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization: Geneve

Alexandersson H, Moberg A. 1997. Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. Internacional Journal of Climatology 17: 25–34.

Brunet M, Bañóon M, García F, Aguilar E, Saladié O, Sigró J, Asín J, López D. 2004. Una aproximación experimental tendente a la minimización del sesgo artificial asociado al tipo de garit meteorológica mediante la observación dual de la temperatura del aire. La Meteorología y el Clima Atlánticos. Publicaciones de la Asociación Española de Meteorología: Badajoz; 93–103.

Brunet M, Saladié O, Jones Ph, Sigró J, Aguilar, E, Moberg A, Lister L, Walther A, Lopez D, Almarza D. 2006: The development of a new dataset of Spanish Daily

Adjusted Temperature Series (SDATS) (1850-2003), International Journal of Climatology, in print, DOI: 10.1002/joc
Della-Marta P, Wanner H. 2006. A method of homogenizing the extremes and mean of daily temperature measurements. Journal of Climate In print.
Moberg, A., P.D. Jones, D. Lister, A. Walther, M. Brunet, J. Jacobeit, L.V. Alexander, P.M. Della-Marta, J. Luterbacher, P. Yiou, D. Chen, A.M.G. Klein Tank, O. Saladie, J. Sigro, E. Aguilar, H. Alexandersson, C. Almarza, I. Auer, M. Barriendos, M. Bergert, H. Bergstrom, R. Bohm, J. Butler, J. Caesar, A. Drebs, D. Founda, F-W. Gerstengarbe, G. Micela, M. Maugeri, H. Osterle, K. Panzic, M. Petrakis, L. Srnec, R. Tolasz, H. Tuomenvirta, P.C. Werner, H. Linderholm, A. Philipp, H. Wanner and E. Xoplaki, 2006, Indices for daily temperature and precipitation extremes in Europe analysed for the period 1901-2000, J. Geophys. Res – Atmospheres (accepted)
Vincent LA, Zhang X, Bonsal BR, Hogg WD. 2002. Homogenization of daily temperatures over Canada. Journal of Climate 15: 1322–1334.

HOMOGENIZATION OF DAILY AIR PRESSURE AND TEMPERATURE SERIES FOR BRNO (CZECH REPUBLIC) IN THE PERIOD 1848–2005

P. Štěpánek[1], L. Řezníčková[2], R. Brázdil[2]

[1] Czech Hydrometeorological Institute, Regional Office Brno, Czech Republic
[2] Institute of Geography, Masaryk University, Brno, Czech Republic
petr.stepanek@chmi.cz

ABSTRACT

Homogenization of daily meteorological series is a difficult task. Several kinds of problem have to be taken into consideration in the course of homogenization: selection of a proper homogenization method with regard to the data used, creation of reference series, completion of missing values, annual course of adjustments, and others. This paper presents an attempt to create a homogeneous series of daily air pressure and temperature readings in the city of Brno (Czech Republic). Two basic approaches were adopted: (i) homogenization of monthly series and projection of estimated smoothed monthly adjustments in annual variation of daily adjustments and (ii) homogenization of daily values in individual months and direct estimation of daily adjustments, again smoothed by low-pass filter. Differences in the results obtained from these two approaches are further discussed.

INTRODUCTION

In the recent years considerably more attention has been devoted to the analysis of the daily data widely recorded and stored in databases. Prior to analysis, the need to homogenize the data and check their quality arises. There is no widely accepted homogenization approach that could be generalized and applied to various meteorological elements, different climatic patterns, etc., and this will probably never be possible. This is, for example, due to the fact that the statistical properties of daily data and regional differences between them make general homogenization of daily values difficult, as well as involving more demanding data handling. During data processing, several kinds of problem have to be taken into consideration. These involve selection of a proper method for homogenization with regard to the data used, i.e. fulfilling all the conditions necessary to applying selected tests of relative homogeneity (e.g. normal distribution), creation of reference series (defining selection criteria), completion of missing values, annual course of adjustments, and others.

Only a few studies, in comparison with monthly or annual data series, have been devoted to techniques addressing daily values. For example, Brandsma (2000) compared monthly adjustments, daily adjustments derived from monthly adjustments (using iterative cubic spline interpolation to preserve monthly adjustments) and daily adjustments derived from weather types. Wijngaard et al. (2003) did not use measured values, but their characteristics, such as diurnal temperature range and its annual mean, as well as the annual mean of the absolute day-to-day differences for temperature, and the annual number of wet days for precipitation. Following various homogeneity tests, these series were labelled as recommendations for further analysis.

Mekis and Vincent (2004) derived daily adjustments from monthly adjustments. These were obtained using linear interpolation between mid-month “target” values objectively chosen so that the average of the daily adjustments over a given month is equal to the monthly adjustment. This approach does not require the creation of a daily reference series or the identification of inhomogeneities in daily temperatures. Moreover, finally homogenized series of daily temperatures are compatible with homogenized monthly data-sets.

The present paper is dedicated to the search for a proper methodology for daily data-set handling and its subsequent application to daily air pressure and temperature series for Brno in the period 1848–2005, with the aim of creating a homogeneous series for Brno with regard to both elements. Although there are, in general, no gaps in the Brno measurements, data are unfortunately not available from a single site, so it becomes necessary to combine different series to get one Brno series suitable for further analysis. The basic Brno stations were tested separately for relative homogeneity and, after homogenization, they were combined using overlap periods. All calculation was performed using AnClim and ProClimDB softwares (Štěpánek, 2006a, 2006b).

1. A BRIEF HISTORY OF METEOROLOGICAL OBSERVATIONS IN BRNO

Meteorological observations in Brno began in 1799, the work of Captain Emeritus Ferdinand Knittelmayer, but his observations for the period 1799–1812 are preserved only in the daily averages. For the subsequent years 1813–1819, the observations exist only in the form of monthly averages. On the basis of several daily readings, meteorological observations were published in the daily newspaper “Mährisch-Ständische Brünner Zeitung” from January 1820 to December 1847. In some years, parallel observations from two places in Brno were also made. Although monthly value series for air pressure, air temperature and precipitation totals have been homogenised and analysed (Brázdil et al., 2005), work with daily readings or daily averages requires further research. For this reason, the analysis provided in this paper works only with data from 1848 onwards.

Meteorological observations after 1848 come from Dr. Paul Olexik (1800–1878), a physician from St. Anne’s hospital (Fig. 1). He was probably making observations from as early as the end of 1845, but it was only from 1848 that his measurements started to be published regularly in the Austrian Meteorological Yearbooks, i.e. when his station became part of the network of the Central Meteorological Institute in Vienna. He observed at 0600, 1400 and 2200 hours. On 3 December 1853 he moved the point of his meteorological observations from the hospital (204 m.a.s.l.) a short distance, to his new flat at Pekařská Street 100 (219 m.a.s.l.). Meteorological observations at this new site continued until 30 June 1878. By this time, Gregor Johan Mendel (1822–1884), abbot of the Augustinian Monastery and a pioneer geneticist, was helping to complement Olexik’s measurements, something he continued alone from 1 July 1878 in the monastery garden (204 m a.s.l.) until 30 November 1883. He began with standard readings at 0700, 1400 and 2100 hours. Alfred Lorenz (1825–1890), a professor at the I. R. Technical University, continued meteorological observations in Brno from the university building (225 m a.s.l.), located close to the city centre, from 1 January 1884 until his death in June 1890. Upon his death, air temperature and pressure measurements definitely stopped and no new place for observation was to be found (Brázdil, 1979).

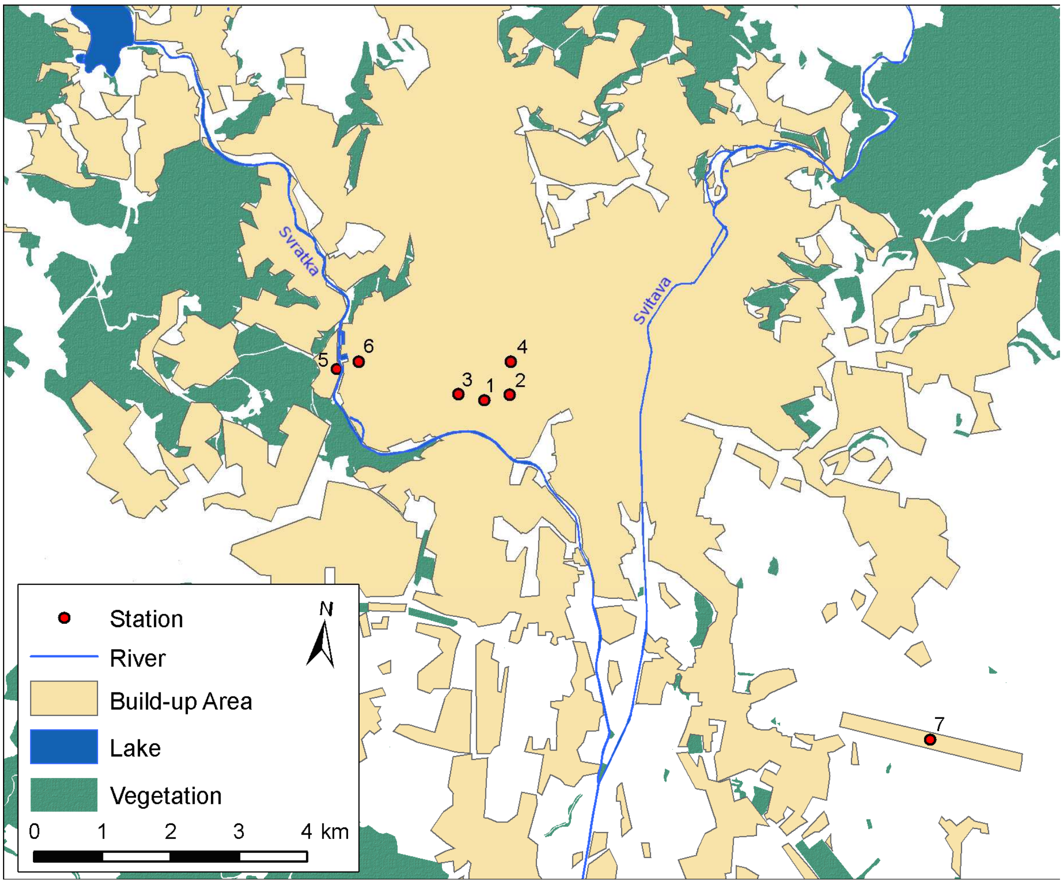


Fig. 1. Location of meteorological stations in Brno: 1 – St. Anne's hospital; 2 – Pekařská Street 100; 3 – Augustinian monastery; 4 – I. R. Technical University; 5 – Pisárky, waterworks; 6 – Květná Street; 7 – Tuřany Airport

However, from 1 June 1890 a meteorological station at the city waterworks in Pisárky (204 m a.s.l.) (further as Brno-Pisárky) began operations, with a full observation programme up to 1937 and with air temperature measurements continuing up to 1962. Further meteorological stations in different parts of Brno were established later, of which only the two used in this paper are mentioned. The first of them was located close to the previous station on Květná Street (further as Brno-Květná), in the garden of the research agricultural institute (223 m.a.s.l.), with observations from 1 August 1922 to 31 December 1970. The second station (Brno-Tuřany) is located at the Brno airport, south-east of the city of Brno (238 m.a.s.l.), i.e. opposite to all the previously mentioned stations, which are concentrated in its western part. Observations started there on 14 April 1958. For this reason, compilation of the Brno daily temperature and pressure series is made with respect to this station.

In summary, addressing knowledge of the history of temperature and pressure measurements in Brno from 1848 onwards, with respect to homogenization, it should be stressed that measurements

- were provided from different parts of Brno, at different altitudes
- were provided by different types of instruments
- were provided in different observation terms before and after 1878
- are limited by lack of available overlap for observations predating 1890.

2. DATA USED

For outlier identification as well as for relative homogeneity testing, other stations with long-term series in the broad surroundings of Brno were also used (Fig. 2). A list of them, with basic characteristics, is given in Table 1. We have used all the measurements, i.e. not only daily averages but also separate series from individual observation hours. As has already been mentioned, as well as standard observations times at 0700, 1400 and 2100 hours local mean time, observations were also carried out at 0600, 1300 and 2200 hours. Finally it was decided, that all the terms should further be treated as if they were 0700, 1400 and 2100 hours in the hope of disclosing possible inhomogeneities arising out of various observing times during homogeneity testing. The original observing hours were taken into consideration during decision-making about adjustments of inhomogeneities found.

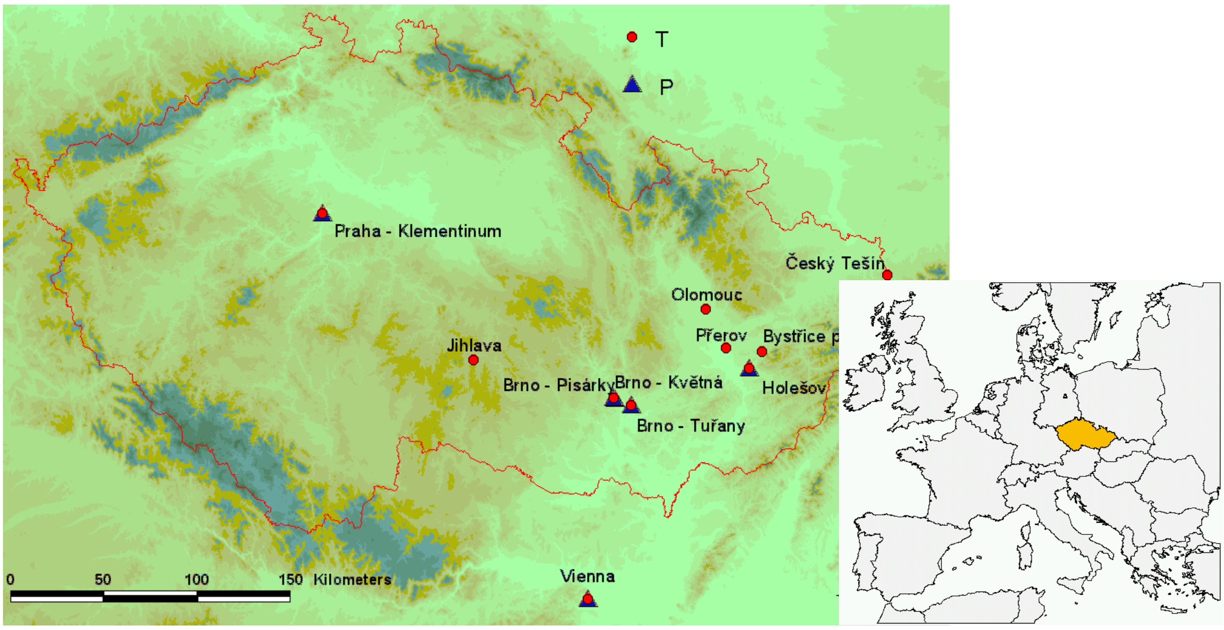


Fig. 2. Geographical distribution of stations used for homogenization of the Brno series (T – air temperature, P – air pressure)

Table 1. Basic information about stations used for homogenization of the Brno series (station coordinates are given for their last or recent locations)

Air temperature

Station name	Latitude (N)	Longitude (E)	Altitude (m.a.s.l.)	Beginning	End	Observing hours
Brno (various places)	49°12´	16°37´	225	1 Jan. 1848	31 Dec. 1889	07 (06), 14 (13), 21 (22)
Brno-Pisárky	49°12´	16°34´	203	1 June 1890	31 May 1962	07, 14, 21
Brno-Květná	49°12´	16°34´	223	1 Aug. 1922	31 Mar. 1970	07, 14, 21
Brno-Tuřany	49°09´	16°˘42´	241	14 Apr. 1958	31 Dec. 2005	07, 14, 21
Bystřice pod Hostýnem	49°24´	17°40´	315	1 Sep. 1865	31 Dec. 2005	07 (06), 14, 21 (22)
Český Těšín	49°44´	18°37´	280	1 Jan. 1885	31 Oct. 1938	07, 14, 21
Holešov	49°19´	17°34´	224	1 July 1895	31 Dec. 2005	07, 14, 21 (22)
Jihlava	49°23´	15°32´	560	27 July 1873	31 Dec. 1934	07 (08), 14, 21 (22)
Olomouc	49°36´	17°15´	215	1 Jan. 1876	31 Dec. 1960	07 (08), 14, 21 (20)
Prague-Klementinum	50°05´	14°25´	191	1 Jan. 1775	31 Dec. 2005	07, 14, 21
Přerov	49°25´	17°24´	203	1 Apr. 1874	31 Dec. 1979	07, 14, 21
Vienna-Hohe Warte	48°13´	16°21´	199	1 Jan. 1872	31 Dec. 2005	07, 14, 19

Air pressure

Station name	Latitude (N)	Longitude (E)	Altitude (m.a.s.l.)	Beginning	End	Observing hours
Brno (various places)	49°12´	16°37´	225	1 Jan. 1848	31 Dec. 1889	07 (06), 14 (13), 21 (22)
Brno-Pisárky	49°12´	16°34´	203	1 June 1890	31 Dec. 1937	07, 14, 21
Brno-Květná	49°12´	16°34´	223	1 Aug. 1922	31 Dec. 1962	07, 14, 21
Brno-Tuřany	49°09´	16°42´	241	14 Apr. 1958	31 Dec. 2005	07, 14, 21
Holešov	49°19´	17°34´	224	1 Jan. 1961	31 Dec. 2005	07, 14, 21
Prague-Klementinum	50°05´	14°25´	191	1 Aug. 1787	31 Jan. 2002	14
Vienna-Hohe Warte	48°13´	16°21´	199	1 Jan. 1872	31 Dec. 2005	07, 14, 19

The correlation coefficients for both elements analyzed are high enough for all stations involved (Fig. 3). Their values were calculated from original data (not from series of first differences), so they are biased by inhomogeneities in a shift (the values would otherwise be higher) and also by trends (the values would be lower if the trend were removed from the series). Correlations of monthly averages are higher than those of daily averages during the winter months, while the opposite holds in summer, i.e. the correlations of monthly averages drop below the values of daily data. From this it follows that both monthly and daily data should be used for data homogenization; daily data are more sensitive to inhomogeneity detection, especially during the summer months.

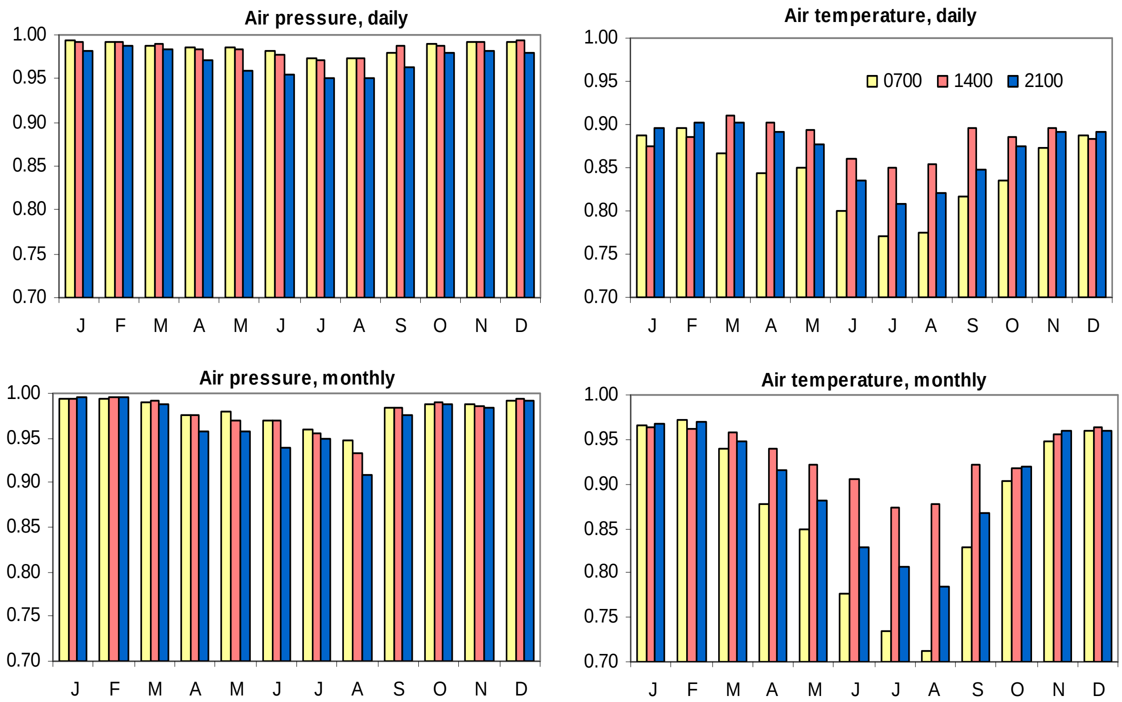


Fig. 3. Medians of correlation coefficients for all pairs of stations, for daily and monthly air temperature (50 values) and air pressure (6 values – without Prague-Klementinum at 1400 hours)

3. HOMOGENIZATION

Homogenization includes the following steps: detection, verification and possible correction of outliers (extreme values), creation of reference series, homogeneity testing (various homogeneity tests), determination of inhomogeneities in the light of test results and metadata, adjustment of inhomogeneities and filling in missing values (Fig. 4).

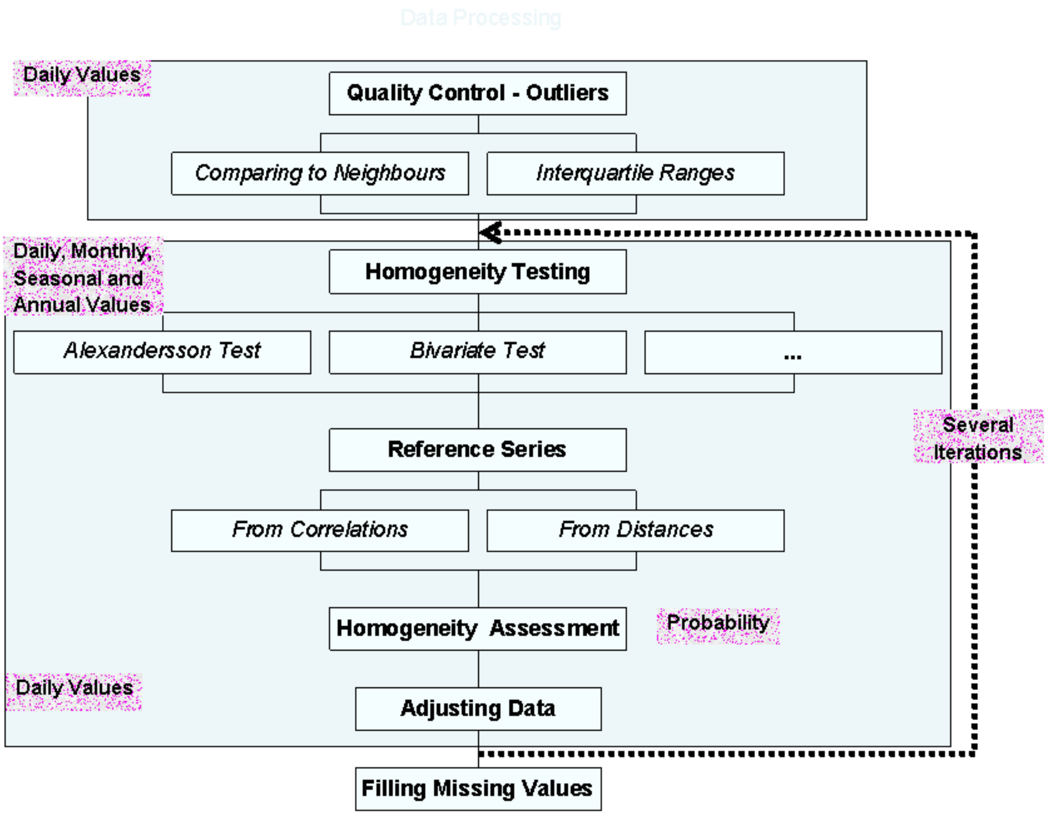


Fig. 4. Plan of the homogenization process

3.1 Outlier identification

Data quality control was carried out in two ways in this study: (i) by applying limits derived from interquartile ranges (either to individual series, i.e. absolutely or, better, to difference series between candidate and reference series, i.e. relatively), (ii) by comparing candidate station values to values from neighbour stations.

In comparisons with neighbour stations, the five best correlated series were selected (correlations calculated from series of first differences – see e.g. Peterson, 1998), the values of correlation coefficients being at least 0.50; no limit for distance or altitude difference has been applied. Only series with the same observation hours were selected. For the evaluation of outliers, various characteristics were considered. A count of statistically significant different neighbours (compared to candidate station) exceeding the confidence limit (0.95) was evaluated by means of difference series (neighbour minus candidate station), for each month individually. Cases in which more than 75% of neighbours differed significantly from the base station values were checked visually. To help in establishing the nature of the outliers, the values of neighbours were standardized with respect to candidate station average and standard deviation and a new (theoretical) value for the candidate station was also calculated – as a weighted average from the standardized values of the neighbours. Further, the coefficient of interquartile ranges (q75–q25) above q75 (or below q25) were evaluated (calculated from the standardized neighbour values), and applied to candidate station value. The reason for this was to assess similarity of neighbour values used with regard to test value: the more values of neighbours are similar, the higher is the value of the coefficient.

The final decision on removing outliers was based on the percentage of the count of significantly different neighbours, difference from "expected value", coefficient of interquartile range, and finally by visual (subjective) comparison of the standardized values of neighbours with the candidate station values. Fig. 5 shows an example of the output for decision-making about outliers.

Id	Year	Month	Day	St_base	Remark	St_1	St_2	St_3	St_4	St_5	Exc_	Exc_	St_1_ba	St_2_ba	St_3_b	St_4_	St_5_	Expect_val
B1BYSH01_T_07:00				315.0	Altitudes,limits	282.0	203.0	560.0	215.0	199.0								
O1TESI01_T_07:00					st_1, distance	111.7												
O3PRER01_T_07:00					st_2, distance		29.0											
B2JIHL01_T_07:00					st_3, distance			237.1										
O2OLK_01_T_07:00					st_4, distance				50.8									
B6VHOH01_T_07:00					st_5, distance					192.6								
B1BYSH01_T_07:00	1886	1	1	-13.6		-6.7	-12.8	-11.5	-10.0	-11.1	1	20.0	-6.9	-12.9	-11.3	-9.9	-13.1	-11.9
B1BYSH01_T_07:00	1886	1	2	-7.4		-2.4	-7.2	-1.0	-6.8	-10.6	1	20.0	-3.0	-7.3	-0.5	-6.7	-12.5	-7.0
B1BYSH01_T_07:00	1886	1	3	1.6		2.6	1.0	1.8	1.9	4.7			1.4	0.9	2.4	2.0	3.8	1.2
B1BYSH01_T_07:00	1886	1	4	2.0		3.4	0.9	1.5	2.2	1.1			2.2	0.8	2.1	2.3	-0.1	1.2
B1BYSH01_T_07:00	1886	1	5	4.4		4.4	3.7	4.5	2.0	0.1	1	20.0	3.1	3.6	5.2	2.1	-1.1	3.2
B1BYSH01_T_07:00	1886	1	6	1.6		2.6	0.6	0.9	0.6	3.7			1.4	0.5	1.4	0.7	2.7	0.6

Fig. 5. Example of output with auxiliary characteristics for quality control evaluation

In some cases, in which at least two neighbours were not available, interquartile ranges for each individual month of the candidate series were applied (i.e. absolutely) and the errors emerging were checked. This method has considerably inferior results in comparison with the relative method, but no other possibility existed for cases in the distant past.

3.2 Homogeneity test

As well as monthly, seasonal and annual averages, series of daily data were also tested. In this case we used all days of a particular month and further an aggregation of "seasons and year" calculated from the first days of all months, the second days, etc. (see Fig. 6). Although such "aggregate" series cannot be used for common time series analysis because the time is "cracked", it can be very useful for the purposes of finding discontinuity (seasonal to annual resolution), while original daily values, even when used only within particular months, can suffer from annual course (this is the case for air temperature rather than air pressure, mainly in winter) and normality is sometimes on the border of the 0.05 significance level. Using the aggregates over seasons and year leads to series for which normality is fulfilled without problems, and thanks to lower signal-to-noise ratio this approach is better for detecting real inhomogeneities in the series. Significant autocorrelations within a number of first lags (days) appear to present a larger problem and have to be further investigated. Series are more persistent in winter with stronger circulation effects, rather than in summer with its prevailing radiation factors.

Data of B1BYSH01_T_07_00_3.txt

Year	Day	XII,I-II	III-V	VI-VIII	IX-XI	I-XII
1924	27	-6.8	11.5	13.3	9.9	7.1
1924	28	-3.4	10.2	14.9	6.0	6.9
1924	29	-4.1	11.0	13.9	6.2	7.0
1924	30	missing	9.2	16.5	6.5	missing
1924	31	missing	missing	missing	missing	missing
1925	1	1.2	6.9	15.9	10.9	8.6
1925	2	-0.0	6.2	16.0	11.1	8.2

Fig. 6. An example of using daily data for homogeneity testing

Several relative homogeneity tests (significance level 0.05) were used: the Alexandersson Standard Normal Homogeneity Test SNHT (Alexandersson, 1986, 1995), the Maronna and Yohai bivariate test (Potter, 1981), the Pettit test (Pettit, 1979), the t-test (Mitchell et al., 1966) and the Easterling and Peterson test (Easterling, Peterson, 1995). Tests were applied to 40-year sections of the series tested for monthly averages and 30-years series of daily data because the alternative hypothesis of the Alexandersson and bivariate tests assumes the presence of only one inhomogeneity in a series (we applied SNHT for a single shift). Series longer than 40 years were divided into several parts with an overlap of ten years (or five years for daily data). This is important in the light of tendencies to overestimation of detected inhomogeneities near the ends of series (see Alexandersson, 1995). Reference series were created separately with respect to each 40-year (30-year) parts of a candidate series (this means with its own selection of neighbours in each part). For daily data, 185 sections of series (of 49 original elements-terms-stations) were created and tested.

The use of series with durations of 40 and/or 30 years seems to be reasonable for homogeneity testing. Shorter series would not be so suitable from a statistical point of view, while, on the other hand, longer series usually contain more than one inhomogeneity (the typical duration of a period with one inhomogeneity does not usually exceed 30–40 years – see e.g. Auer et al., 2001).

To ensure that only one inhomogeneity detected by the Alexandersson or bivariate tests was present in a series, a further modification was introduced into the AnClim software. The series was divided at the position of a detected inhomogeneity and sections before and after it were tested separately. If no other inhomogeneity was found, we can rely on the results of the given test for the whole length of the series (especially the significance of a test statistic).

3.3 Reference series creation

Reference series were created in two ways: (i) an average from the best correlated stations, (ii) an average from nearest stations. Correlation coefficients used for station selection were calculated from the series of first differences, when inhomogeneities are manifested in the only value (see e.g. Alexandersson, Moberg, 1996; Peterson, 1998). Various types of reference series with analysis of their advantages and drawbacks have been discussed, for example, by Štěpánek (2005).

The values of correlation coefficients were not allowed to drop below 0.60 between neighbour stations (selection by means of correlation) and no distance or altitude limits were applied as additional conditions for air pressure and temperature. Weighted averages were calculated using correlations and/or reciprocal values of station distances as weights. Values of selected neighbour stations were standardized to candidate station average and standard deviation to avoid problems with biased reference series. This can often happen in the event of missing data in one of the neighbour series. The standardization was done for each particular month individually (also for daily data). No transformation of values has been applied to the data.

In the first stage, a list of proposed neighbour stations was obtained, which was subsequently checked and its approved version was then finally used for the reference series calculation.

3.4 Assessment of detected inhomogeneities

The main criterion for determining a year of inhomogeneity was the probability of the given inhomogeneity, i.e. the ratio between the count of detections for a given year from all tests for the given station (using all types of reference series, tests, daily, monthly, seasonal and annual series) and the count of all theoretically possible detections. The count of detections for groups of years was also taken into account (some inhomogeneities started in the course of the year and thus were manifested in at least two years). If metadata did not confirm the detected shift (in most cases), the percentage limit of all possible detections was taken higher and some other information (e.g. distribution of the given year within individual months or seasons, graphs of differences with reference series and some other characteristics) was required to decide whether the undocumented inhomogeneity could be regarded as "indubitably" proven and consequently corrected. For assessment of the inhomogeneities detected, the Real Precision Index (RPI – see Petrovic, 2004) may also be applied to find sections of series that exhibit change in the quality of measurements.

3.5 Adjustment of inhomogeneities

Adjustment of inhomogeneities detected was addressed by means of the reference series calculated from the average over the five stations with the highest correlation coefficients with the series being adjusted (correlations were calculated again from the series of first differences). The adjustment value was estimated as the difference between averages calculated from difference series between the tested and the reference series. The start of inhomogeneity was allocated to a particular month (where this was possible).

When dealing with daily data, there are several approaches to adjusting data for inhomogeneities detected. We may use either monthly adjustments which can be distributed into individual days (e.g., Mekis and Vincent, 2004) or we can calculate adjustments for daily data directly.

Using monthly data in this paper, the estimated individual monthly adjustments were smoothed by low-pass filter (weights applied to adjacent months were approximately 1, 2, 1) to suppress the influence of random errors in the series (the effect of smoothing results in a more realistic annual course for the adjustments, in line with what is better physically justified). The monthly adjustments obtained were then distributed (interpolated) among individual days and the final daily adjustments (again possibly smoothed to eliminate the edges of lines occurring each month) were then applied to data.

In the second case, the daily adjustments difference series (reference and tested) for each day of the year were used, taking 20 years before and after the change. Final daily adjustments were then smoothed using a low-pass filter for 60 days (to each side).

Various characteristics were analyzed before applying the adjustments: increment of correlation coefficients between candidate and reference series after adjustments, change of standard deviation in differences before and after the change, presence of linear trend, etc. In the event of any doubts, the adjustments were not applied.

3.6 Further considerations

The above-mentioned steps were performed in several iterations. At each iteration, more precise results were obtained. Missing values were filled in only after homogenization and adjustment of inhomogeneities in the series. The reason for this was that the new values were estimated from data not influenced by possible shifts in the series. Moreover, when missing data are filled in before homogenization, they may influence inhomogeneity detection in a negative way. The gaps were filled by means of linear regression between filled value series (dependent variable) and a reference series (independent variable), separately for each month. For assessing the quality of the process, various statistics were monitored, e.g. differences of averages and standard deviations in periods before and after the gap.

4. HOMOGENIZATION RESULTS

As has been shown above, the values of correlation coefficients for daily data (using each month individually) are comparable with values gained from monthly averages. The same holds true of correlations between tested and reference series. The medians of correlation coefficients for monthly air temperature range from 0.87 in the summer months to 0.98 in the winter months for individual observation hours; again the results at 1400 hours correlate the best. For daily data, the correlations for individual months range from 0.87 to 0.95. For air pressure, daily data correlates between 0.97 in summer and 0.99 in winter, monthly data between 0.94 and 0.99.

From these results, it follows that it is worth working with daily data in the course of homogenization, even if it is more demanding compared to “simple” monthly averages. By employing daily data we have longer series (28 to 31 times, depending on number of days in a particular month) and we can then also better detect shifts near the end of the series (not resolvable for monthly averages with breaks of less then five years to the end of series). Fig. 7 gives count of inhomogeneities detected for daily and monthly series by the Alexandersson test with reference series created by means of correlations.

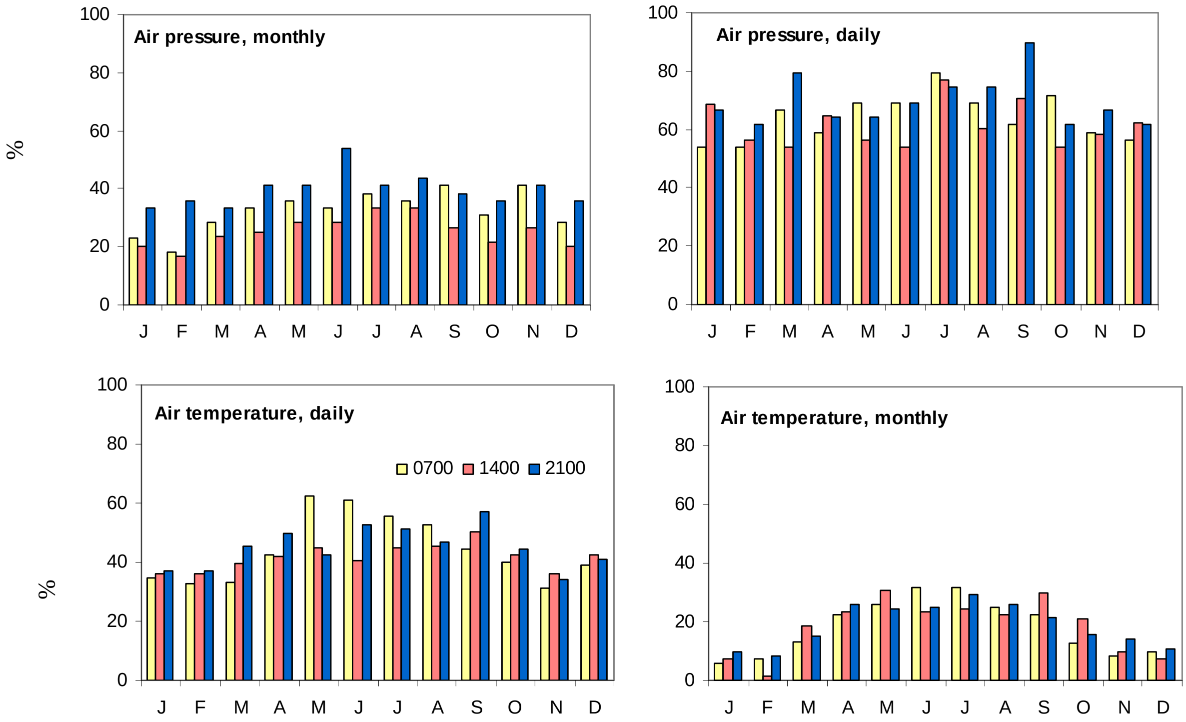


Fig. 7. Percentage of inhomogenities in air pressure and temperature series for daily data and monthly averages, detected by the Alexandersson test, related to the total number of series used

The annual course of numbers of inhomogeneities is evident from the figure, as are the differences between air pressure and temperature readings, as well as observation hours. The large difference between monthly- and daily-based detections is, among other things, due to the fact that in the Alexandersson test the series is divided into sections in the position of each detected break. Since the series contain more members, we are able to detect relatively more inhomogeneities (mainly in the shorter sections). In this sense, the numbers between daily and monthly series are not comparable. But the aim was to show that during homogenization we should try to use information that is as dense as possible, using daily data, individual observation hours, etc.

The advantages of using daily data mentioned above are apparent from the example in Fig. 8. In the event of missing values and breaks near the ends of series it is more difficult to detect inhomogeneities in the series if one works with only monthly data.

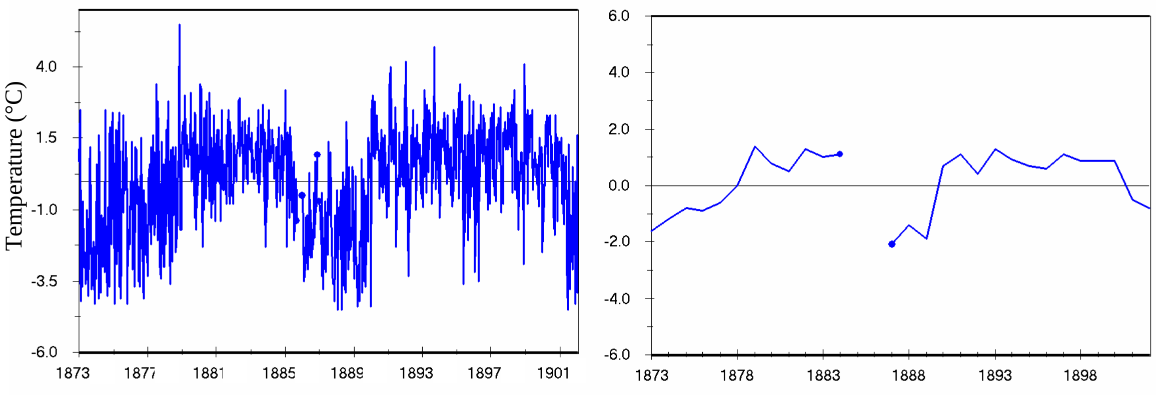


Fig. 8. Differences between tested and reference series for daily (left) and monthly (right) data for Brno, air temperature at 0700 (0600) hours, July 1873–1902

4.1 Homogenized pressure and temperature series of Brno

The creation of homogenized air pressure and temperature series for Brno covering the period 1848–2005 consists of several steps. First, the individual series for the different Brno stations (Brno stations before 1890 & Brno-Pisárky, Brno-Května, Brno-Tuřany – see Chapter 2) were homogenized according to the methodology described in Chapter 4. In the second step, a common compiled Brno series was developed by adjusting the individual parts. Starting from the recent observing station at Brno-Tuřany (reference station, 1958–2005), Brno-Květná data were adjusted to its measurements to obtain a series for the period 1923–2005. In the next step, the Brno-Pisárky station was adjusted to the combined Květná-Tuřany reference series to obtain a series for the period 1890–2005 (Fig. 9). This approach was applied separately for each observation time (0700, 1400 and 2100 hours).

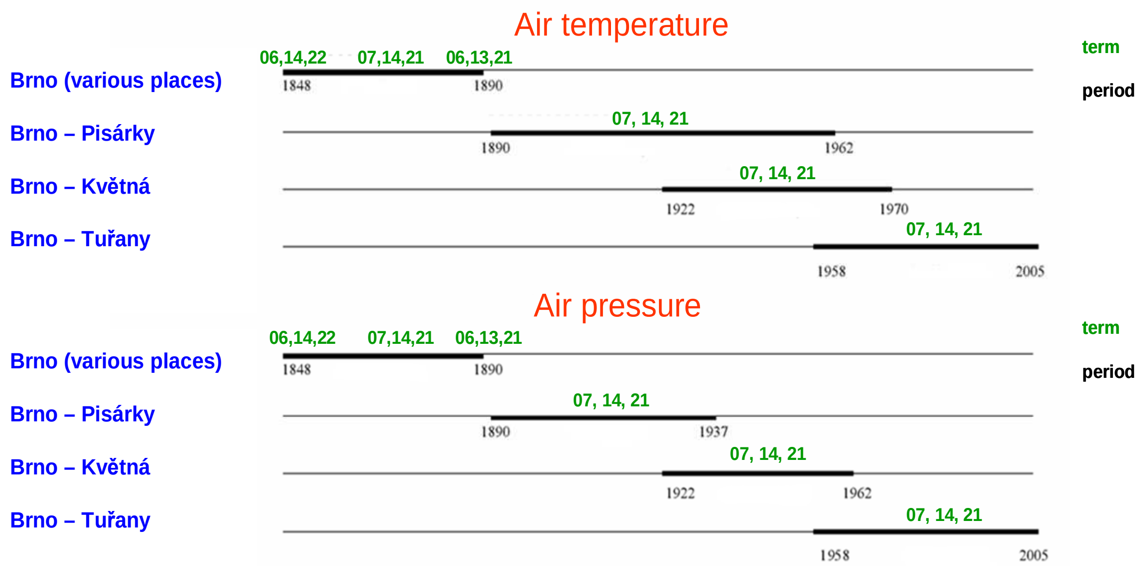


Fig. 9. Scheme of creation for the series compiled by combining measurements from several locations in Brno

The principle of combination used for the individual Brno stations is identical to that employed for adjustments of inhomogeneities applied to data in the course of homogenization. Two approaches may be selected: one that uses monthly averages or one that works directly from daily data. The only difference is that final offsets are not computed by comparing periods before and after the change; in this case we use the whole period in common (shortened to 20 years if it is longer). The overlap periods vary from 5 years (air pressure) or 13 years (air temperature) in the first round to 15 years (air pressure) or 20 years (air temperature) in the second round.

Fig. 10 gives an example of when final adjustment is obtained either from monthly averages or through direct use of daily data. It seems appropriate to calculate adjustments from daily values using a low-pass filter for 60 days, or, leading to the same results, using a low-pass filter for two months and subsequently distributing the smoothed monthly adjustments into daily values.

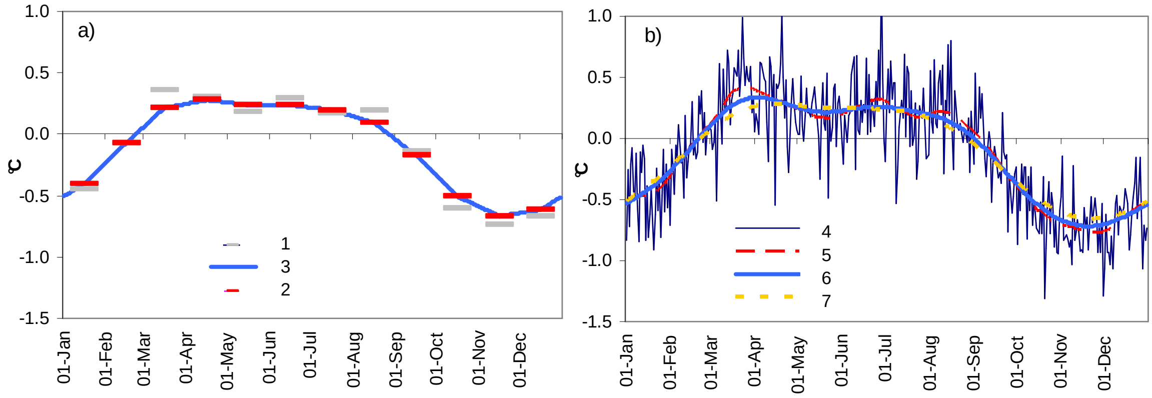


Fig. 10. Annual variations of adjustments applied to air temperature series at 1400 hours for Brno-Květná to the reference station Brno–Tuřany: a) monthly-based approach (1 – raw adjustments, 2 – smoothed adjustments, 3 – smoothed adjustments distributed into individual days), b) daily-based approach (4 – individual calendar day adjustments, 5 – daily adjustments smoothed by low-pass filter for 30 days, 6 – for 60 days, 7 – for 90 days)

The values measured at different observation hours exhibited quite different annual variations of adjustment, making it useful to work with them directly, and not just with calculated daily averages. For example, depending on the formula used for the calculation of daily averages, real inhomogeneities may be masked there.

A fully compiled series for the period 1848–2005 was again tested for homogeneity as a whole. Finally, homogenous Brno pressure and temperature series for 1848–2005 at 0700, 1400 and 2100 hours were obtained, from which corresponding daily and monthly averages were calculated. Fig. 11 shows fluctuations in annual averages both for the compiled homogeneous Brno series and the original series from the various places, in which it was derived.

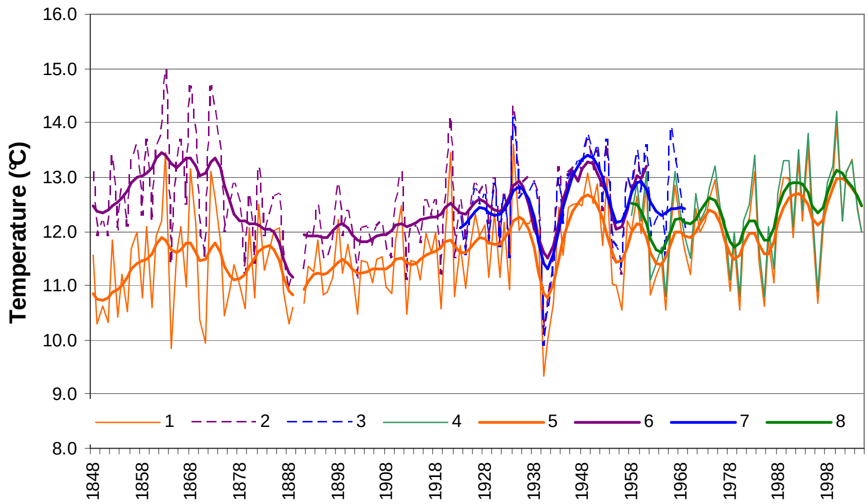


Fig. 11. Fluctuations of annual averages of air temperature series at 1400 hours (1 – compiled Brno series, 2 – Brno–various places and Brno-Pisárky, 3 – Brno-Květná, 4 – original Brno–Tuřany, 5, 6, 7, 8 –series smoothed by Gaussian low pass filter for 10 years)

5. CONCLUSIONS

This work was carried out in quest of a proper methodology for daily data homogenization and made an attempt to apply it subsequently to daily air pressure and temperature series for Brno in the period 1848–2005. Different methods for the homogenization of daily values were sought, and finally applied to find possible inhomogeneities and to obtain adjusted, homogeneous series. Although further investigation in this matter is required, progress so far may be summarized as follows:

(i) Two basic approaches, based on the homogenization of monthly series and projection of estimated monthly adjustments into a smoothed annual course of daily adjustments, or homogenization of daily values of individual months, estimating proper adjustments for each calendar day with smoothing adjustments, can be used.

(ii) The same final adjustments may be obtained from either monthly averages or through direct use of daily data. For the daily-values-based approach, it seems reasonable to smooth them with a low-pass filter for 60 days. The same results may be derived using a low-pass filter for two months (weights approximately 1:2:1) and subsequently distributing the smoothed monthly adjustments into daily values.

(iii) The values of the correlation coefficients between the candidate and reference series for daily data (working with each month individually) are comparable with values gained from monthly averages, although daily data are better in some months, monthly data in others. For this reason, a combination of both approaches in (i) is useful.

(iv) It is profitable to analyze series of individual observation hours because inhomogeneities manifest in different ways within their series – this is the case for the number of inhomogeneities detected, the value of change, the correlations between reference and tested series (and thus detectability of inhomogeneities) and other characteristics. Series of daily averages can serve as complementary information in the course of homogeneity test evaluation. For inhomogeneity assessment, we recommend the use of as much information as possible.

(v) The data processing in this work has been done by means of LoadData software (application for downloading data from central database, e.g. Oracle), ProClimDB software for processing whole datasets (finding outliers, combining series, creating reference series, preparing data for homogeneity testing, etc.) and AnClim software for homogeneity testing (http://www.klimahom.com/software). Further development of the software, e.g. connection with R software, is to be assumed.

Acknowledgements

The authors would like to acknowledge the financial support of the Grant Agency of the Czech Republic for project no. 205/05/0858. Sincere thanks should be extended to Ms Zuzana Pilařová, Institute of Geography, Masaryk University, Brno for digitizing and checking a great part of data used for analysis. Mr. Tony Long, Auchrae, Scotland, undertook the English editing.

REFERENCES

Alexandersson, A. (1986): A homogeneity test applied to precipitation data. Journal of Climatology, 6, 661–675.

Alexandersson, A. (1995): Homogenity testing, multiple breaks and trends. In: Proc. 6th Int. Meeting on Stat. Climatology, Galway, Ireland, 439–441.

Alexandersson, H., Moberg, A. (1996): Homogenization of Swedish temperature data. Part I: A homogeneity test for linear trends. In: Moberg, A: Temperature Variations in Sweden Since the 18th Century. The Department of Physical Geography, Stockholm University. Dissertation Series No. 5, 98 pp.

Auer, I., Böhm, R., Schöner, W. (2001): Austrian long-term climate 1767–2000, Multiple instrumental climate time series form Central Europe. Österreichische Beiträge zu Meteorologie und Geophysik, Heft 25, Wien, 155 pp.

Brandsma, T. (2000): Weather-type dependent homogenization of daily Zwanenburg/De Bilt temperature series. http://www.met.hu/omsz.php?almenu_id=omsz&pid=seminars&pri=6&mpx=1&sm0=0&tfi=brandsma

Brázdil, R. (1979): Historie měření srážek v Brně (History of precipitation measurements in Brno). Scripta Fac. Sci. Nat. UJEP Brunensis, 9, Geographia 2, 55–74.

Brázdil, R., Valášek, H., Macková, J. (2005): Meteorologická pozorování v Brně v první polovině 19. století. Historie počasí a hydrometeorologických extrémů (Meteorological Observations in Brno in the First Half of the Nineteenth Century. History of Weather and Hydrometeorological Extremes). Archiv města Brna, Brno, 452 pp.

Easterling, D. R., Peterson, T. C. (1995): A new method for detecting undocumented discontinuities in climatological time series. International Journal of Climatology, 15, 369–377.

Mekis, É., Vincent, L. (2004): New developments in the homogenization of precipitation and temperature series in Canada. In: Fourth seminar for homogenization and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Geneva, 39–62.

Mitchell, J. M., ed. (1966): Climatic Change. WMO, T.N. 79, Geneva, 80 pp.

Peterson, T. C. (1998): Homogeneity adjustments of in situ atmospheric climate data: a review. International Journal of Climatology, 18, 1493–1517.

Petrovic, P. (2004): Detecting of inhomogeneities in time series using Real Precision Metod. In: Fourth seminar for homogenization and quality control in climatological

databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Geneva, 79–88.
Pettit, A. N. (1979): A non-parametric approach to the change-point problem. App. Statist., 28, 2, 126–135.
Štěpánek, P., 2005: Variabilita teploty vzduchu na území České republiky v období přístrojových měření (Air Temperature Fluctuations in the Czech Republic in the Period of Instrumental Measurements). Ph.D. Thesis. Institute of Geography, Masaryk University, Brno, 136 pp.
Štěpánek, P. (2006a): AnClim - software for time series analysis. Institute of Geography, Masaryk University, Brno, 1.47 MB. http://www.klimahom.com/software/AnClim.html
Štěpánek, P. (2006b): ProClimDB – software for processing climatological datasets. CHMI, regional office Brno. http://www.klimahom.com/software/ProcData.html
Wijngaard, J. B. , Klein Tank, A. M. G. , Können, G. P. (2003): Homogenity of 20th century European daily temperature and precipitation series. Int. J. Climatol., 23**,** 679–692

DEVELOPMENT OF MASH HOMOGENIZATION PROCEDURE FOR DAILY DATA

Tamás Szentimrey
Hungarian Meteorological Service
szentimrey.t@met.hu

1. INTRODUCTION

The MASH procedure was developed originally for homogenization of monthly series. It is a relative method, and depending on the distribution of examined meteorological element, additive (e.g. temperature) or multiplicative (e.g. precipitation) model can be applied. In the earlier program system, MASHv2.03, the following subjects were elaborated for monthly series: series comparison, break point (changepoint) and outlier detection, correction of series, missing data complementing, automatic usage of meta data and last but not least a verification procedure to evaluate the homogenization results.
The new version, MASHv3.01, was developed for homogenization of daily series as well as for quality control of daily data and missing daily data completion. During the procedure normal distribution is assumed, therefore at the present version of the software additive model can be applied, that is appropriate e.g. for temperature elements.

2. RELATION OF DAILY AND MONTHLY HOMOGENIZATION

The alternative possibilities are as follows:

– To use the detected monthly inhomogeneities directly for daily data homogenization.
– Direct methods for daily data homogenization.

The problems connected with the possibilities:

– The direct usage of the detected monthly inhomogeneities is probably not sufficient.
– Direct methods for daily data homogenization is probably not enough efficient thinking of the larger variability (less signal to noise ratio).

So we have the following question:
How can we use the valuable information of detected monthly inhomogeneities for daily data homogenization?

3. THE ADDITIVE MODEL OF RELATIVE METHODS

Relative methods can be applied if there are more station series given which can be compared mutually.

3.1 Additive model for daily values (e.g. temperature)

In case of relative methods, the additive model for more daily series in a small climate region is as follows,

$$X^{st}(y,m,d)=\mu(y,m,d)+E^{st}(m,d)+IH^{st}(y,m,d)+\varepsilon^{st}(y,m,d) \qquad (1)$$

where the notations are, st : station, y : year, m : month, d : day,

furthermore $\mu(y,m,d)$ is the common and unknown climate change signal, $E^{st}(m,d)$ are the spatial expected values, $IH^{st}(y,m,d)$ are the inhomogeneity signals and $\varepsilon^{st}(y,m,d)$ are normal white noise series. As concerns the type of $\mu(y,m,d)$ there is no assumption about the shape of this signal.

3.2 Additive model for monthly means

From daily model (1) can be obtained the following model for the monthly means,

$$X_m^{st}(y)=\mu_m(y)+E_m^{st}+IH_m^{st}(y)+\varepsilon_m^{st}(y) \qquad (2)$$

where the means are, $X_m^{st}(y)=\overline{X^{st}}(y,m)$, $\mu_m(y)=\overline{\mu}(y,m)$, $E_m^{st}(y)=\overline{E^{st}}(y,m)$, $IH_m^{st}(y)=\overline{IH^{st}}(y,m)$, $\varepsilon_m^{st}(y)=\overline{\varepsilon^{st}}(y,m)$,

and $\mu_m(y)$ is the common and unknown monthly climate change signals, E_m^{st} are the monthly spatial expected values, $IH_m^{st}(y)$ are the monthly inhomogeneity signals and $\varepsilon_m^{st}(y)$ are normal white noise series. There is no assumption about the shape of signals $\mu_m(y)$ and the type of inhomogeneity signals is in general a step-like function in time with unknown break points and shifts.

4. USAGE OF ESTIMATED MONTHLY INHOMOGENEITIES FOR DAILY HOMOGENIZATION

4.1 Possibilities and problems

Direct methods for daily data homogenization are probably not enough efficient owing to the larger variability that means less signal to noise ratio. However the direct usage of the detected monthly inhomogeneities for daily homogenization is probably not sufficient.

Let us assume that we have the estimation $I\hat{H}_m^{st}(y)$ with good quality for the monthly mean inhomogeneities $IH_m^{st}(y) = \overline{IH^{st}}(y,m)$. It is valuable information but the direct usage of such step curve for daily homogenization may be problematic, as it can be seen in Fig. 1. Consequently the question is how can we obtain appropriate smooth estimation $I\hat{H}^{st}(y,m,d)$ for daily inhomogeneities $I\hat{H}^{st}(y,m,d)$ by using the estimated monthly inhomogeneities $I\hat{H}_m^{st}(y)$?

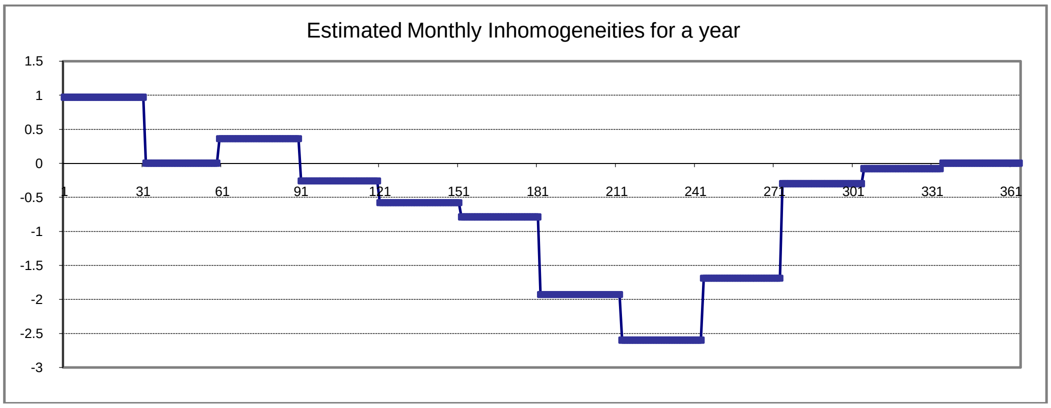


Fig. 1. Example for estimated monthly ihomogeneities for a year

4.1.1 Smoothing according to the method of Vincent et al.

According to Vincent et al. (2002), (Mekis, 2006) the following condition is given for the daily inhomogeneity estimation $I\hat{H}^{st}(y,m,d)$.

The condition for the monthly means is, $\overline{I\hat{H}^{st}}(y,m) = I\hat{H}_m^{st}(y)$.

It seems to be a natural condition if we think on the equality of $\overline{IH^{st}}(y,m) = IH_m^{st}(y)$, however it is possible that too strong inhomogeneities may be obtained occasionally.

4.1.2 Smoothing according to the MASH method

Let us consider another train of thought. Let us not forget that the monthly estimates, $I\hat{H}_m^{st}(y)$, are not real values, these are estimated values only, thus stochastic variables. We have to be aware that to know the real $IH_m^{st}(y)$ is impossible. Consequently the monthly estimates $I\hat{H}_m^{st}(y)$ may be modified, but the modification must be controlled of course.The essence of the applied procedure is as follows.

i, Smooth estimation $I\hat{H}^{st}(y,m,d)$ for daily inhomogeneities by using the monthly estimates $I\hat{H}_m^{st}(y)$ with a not too strong condition, e.g.:

$\exists d_0 : I\hat{H}^{st}(y,m,d_0) = I\hat{H}_m^{st}(y)$

ii, Test of hypothesis to control the new monthly estimations which are the monthly means of daily estimations $I\hat{H}^{st}(y,m,d)$: $I\tilde{H}_m^{st}(y) := \overline{I\hat{H}^{st}(y,m)}$

Fig. 2. is an illustration of smoothing applied at MASH method.

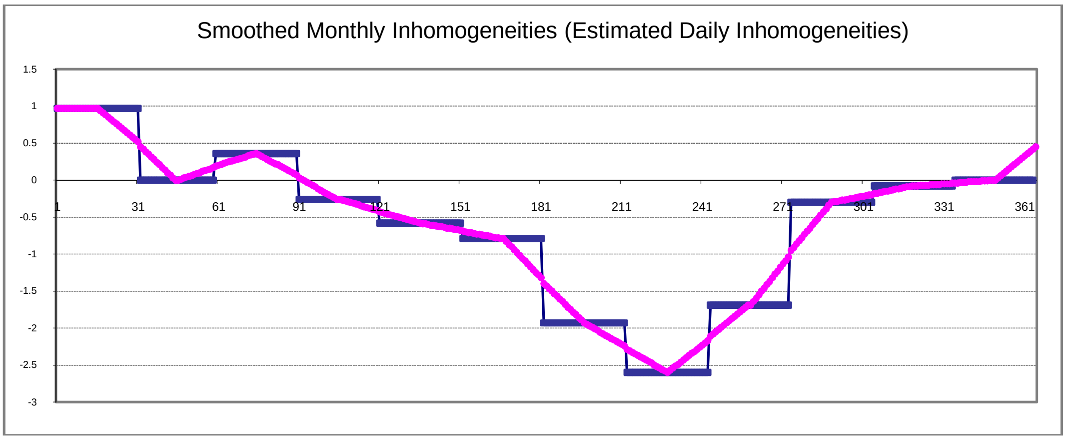


Fig. 2. Example for smoothing of monthly inhomogeneities

4.2 The MASH procedure for daily data homogenization

The algorithm of MASH for daily data homogenization is as follows.

1. Monthly means $X_m^{st}(y)$ from daily data $X^{st}(y,m,d)$.
2. MASH homogenization procedure for monthly series $X_m^{st}(y)$,
   estimation of monthly inhomogeneities: $I\hat{H}_m^{st}(y)$
3. On the basis of estimated monthly inhomogeneities $I\hat{H}_m^{st}(y)$,
   smooth estimation for daily inhomogeneities: $I\hat{H}^{st}(y,m,d)$.
4. Homogenization of daily data:
   $\tilde{X}^{st}(y,m,d) = X^{st}(y,m,d) - I\hat{H}^{st}(y,m,d)$.
5. Quality Control for homogenized daily data $\tilde{X}^{st}(y,m,d)$.
6. Missing daily data complementing.
7. Monthly means $\tilde{X}_m^{st}(y)$ from homogenized, controlled, complemented
   daily data $\tilde{X}^{st}(y,m,d)$.
8. Test of homogeneity for the new monthly series $\tilde{X}_m^{st}(y)$ by MASH.
   Repeating steps 2-8 with $\tilde{X}_m^{st}(y)$, $\tilde{X}^{st}(y,m,d)$ if it is necessary!

The procedure includes also quality control (QC) and missing data completion for the daily data.

5. INTERPOLATION TECHNIQUE USED FOR QUALITY CONTROL (QC) AS WELL AS DATA COMPLEMENTING

In this session the mathematical background of the applied interpolation technique is presented. Let us introduce the following notations.

5.1 Notations

Daily data for a given month:
$X_j(t) \in N(E_j(t), D_j(t)) \quad (j = 1,.., M \text{ station}; t = 1,...,30)$
Candidate data: $X_j(t)$, and reference data: $X_i(t)\ (i \neq j)$.
Interpolation of the candidate data:

$$\hat{X}_j(t) = \lambda_{j0}(t) + \sum_{i \neq j} \lambda_{ji}(t) X_i(t) \quad \text{where} \quad \sum_{i \neq j} \lambda_{ji}(t) = 1 \ .$$

RMS error and representativity values: $RMSE_j(t)$, $REP_j(t) = 1 - \frac{RMSE_j(t)}{D_j(t)}$

The optimum interpolation parameters $\lambda_{j0}^{opt}(t)$, $\lambda_{ji}^{opt}(t)\ (i \neq j; t = 1,..,30)$ which minimize $RMSE_j(t)$ are uniquely determined by the expectations, standard deviations and the correlations. However we have the problem how we can estimate these necessary daily statistical parameters.

5.2 Assumptions for the daily statistical parameters

i, $E_j(t) - E_i(t) = e_{ji}$, $D_j(t) / D_i(t) = d_{ji}$, $(i \neq j; t = 1,..,30)$
ii, $\mathrm{corr}(X_{j_1}(t_1), X_{j_2}(t_2)) = r_{j_1 j_2}^S \cdot r_{t_1 t_2}^T \quad (j_1, j_2 = 1,...M; t_1, t_2 = 1,..,30)$
$r_{j_1 j_2}^S$: correlation structure in space, $r_{t_1 t_2}^T$: correlation structure in time
$\Leftrightarrow$ Partial corr.: $\mathrm{corr}_{X_{j_1}(t_2)}(X_{j_1}(t_1), X_{j_2}(t_2)) = \mathrm{corr}_{X_{j_2}(t_1)}(X_{j_1}(t_1), X_{j_2}(t_2)) = 0$

5.3 Statements

If the assumptions i, ii, are fulfilled then
$\lambda_{j0}^{opt}(t) \equiv \lambda_{j0}^{opt}$, $\lambda_{ji}^{opt}(t) \equiv \lambda_{ji}^{opt}$, $REP_j^{opt}(t) \equiv REP_j^{opt} \quad (t = 1,..,30)$,
where $\lambda_{j0}^{opt}, \lambda_{ji}^{opt}$, REP_j^{opt} are the optimal parameters of monthly interpolation:

$$\hat{\overline{X}}_j(t) = \lambda_{j0} + \sum_{i \neq j} \lambda_{ji} \overline{X}_i \quad \text{where} \quad \sum_{i \neq j} \lambda_{ji} = 1 \ .$$

5.4 Consequences

The monthly statistical parameters can be used for daily interpolation.

i, Data completion based on interpolation: $\hat{X}_j(t) = \lambda_{j0}^{opt} + \sum_{i \neq j} \lambda_{ji}^{opt} X_i(t)$

ii, Quality control can be based on the following standardized error:

$$Z_j(t) = \frac{X_j(t) - \hat{X}_j(t)}{D_j(t)\left(1 - REP_j^{opt}\right)} \in N(0,1)$$

where $\lambda_{j0}^{opt}, \lambda_{ji}^{opt}$, REP_j^{opt} are the optimal parameters of monthly interpolation, and $D_j(t)$ is the daily standard deviation.

6. TEST OF HYPOTHESIS FOR THE STANDARDIZED ERROR SERIES

The standardized errors $Z(t) \in N(0,1)$ $(t = 1,..,n)$ if data have good quality.

But we have the problem: $\mathrm{P}\left(\max_t |Z(t)| < z \right)$ depends on the autocorrelation.

6.1 Statement

i, If $Z(t)$ $(t = 1,..,n)$ is a Markov process, furthemore

ii, $\mathrm{P}\left(|Z(t)| < z \mid |Z(t-1)| < z \right) \geq \mathrm{P}\left(|Z(t)| < z \right) \qquad (t = 2,..,n),$

then $\mathrm{P}\left(\max_t |Z(t)| < z \right) \geq \prod_{t=1}^{n} \mathrm{P}\left(|Z(t)| < z \right).$

Example: If $Z(t)$ $(t = 1,..,n)$ is a normal AR(1) process then i, ii, are fulfilled.

6.2 Decision according to test of hypothesis

The data $Z(t)$ is wrong if and only if $|Z(t)| > z_p$ where critical value z_p is defined by the significance level p (e.g.: p=0.01) as,

$$\mathrm{P}\left(\max_t |Z(t)| < z_p \right) \geq \left(2\Phi(z_p) - 1 \right)^n = 1 - p \quad ,$$

where $\Phi(z)$ is the standard normal distribution function.

6.3 Multiple QC for daily data

More standardized error series are examined without common reference series in order to separate the wrong data of the candidate station. Correction of the wrong data can be based on confidence intervals.

7. EXAMPLE FOR APPLICATION OF MASH FOR DAILY DATA

Examined data: daily temperature series (1901-1930), 10 stations in Hungary.
Temperature element: (max+min)/2. The homogenization procedure 4.2. was implemented and some partial results are presented in Table 1, 2, 3.

Table 1. Partial results of Quality Control for daily data (output ERROR.RES)

Detected errors in September 1903 at Station 10

	st1	st2	st3	st4	st5	st6	st7	st8	st9	st10
1903 9 1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-3.4
1903 9 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-2.2
1903 9 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-3.1
1903 9 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-5.0
1903 9 5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-2.6
1903 9 6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-2.7
1903 9 7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-4.9
1903 9 8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-2.9
1903 910	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.8
1903 911	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-5.5
Original Data										
1903 9 1	20.5	17.1	20.0	17.5	21.0	18.1	18.7	19.5	19.3	12.3
1903 9 2	19.8	17.9	20.8	15.5	21.5	14.9	18.4	19.3	20.2	13.2
1903 9 3	19.1	17.3	20.8	15.5	21.3	15.8	18.5	17.2	17.5	11.3
1903 9 4	19.7	17.5	19.8	15.3	19.0	15.8	18.8	19.2	17.2	10.4
1903 9 5	20.3	17.8	20.5	16.0	21.0	17.4	19.3	20.4	17.8	13.2
1903 9 6	20.9	18.7	21.3	17.3	20.0	18.6	19.9	21.4	18.8	13.8
1903 9 7	22.9	21.5	22.5	17.8	22.0	19.5	18.9	23.6	19.0	13.9
1903 9 8	22.5	20.9	25.0	19.0	23.0	19.8	19.1	23.5	19.5	15.5
1903 910	17.7	18.4	17.0	13.8	13.6	19.0	14.3	18.9	13.7	12.7
1903 911	16.5	13.7	18.3	11.8	14.5	13.5	13.1	18.8	14.1	6.2
Longterm means in September										
	16.7	15.9	16.2	14.9	15.9	15.5	14.6	17.0	14.7	16.6

Table 2. Verification results for the annual series (output MASHVERI.RES)

TEST STATISTICS for ANNUAL SERIES (OUTPUT of MASH)
Critical value (significance level 0.05): 20.53

1. Test Statistics Before Monthly Homogenization

Station	TSBM	Station	TSBM	Station	TSBM
4	317.85	6	241.41	2	155.04
9	127.66	7	91.66	10	68.36
1	62.55	8	61.84	5	42.06
3	15.82	AVERAGE:	118.42		

2. Test Statistics After Monthly Homogenization

Station	TSAM	Station	TSAM	Station	TSAM
7	28.64	5	25.11	9	22.73
4	18.52	1	18.12	8	15.26
6	14.96	2	14.82	10	12.41
3	10.26	AVERAGE:	18.08		

3. Test Statistics After Monthly&Daily Homogenization

Station	TSAMD	Station	TSAMD	Station	TSAMD
7	28.89	5	25.40	2	25.06
9	21.98	1	17.60	4	16.52
8	15.23	6	14.66	3	9.69
10	9.00	AVERAGE:	18.40		

Table 3. Average of verification results for the monthly series

```
AVERAGED TEST STATISTICS FOR MONTHLY SERIES (10 Stations)
Average of Test Statistics Before Monthly Homogenization: TSBM
Average of Test Statistics After Monthly Homogenization: TSAM
Average of Test Statistics After MonthlyDaily Homogenization:TSAMD
MONTH       TSBM       TSAM       TSAMD
    1       28.5       12.0        12.1
    2       21.1       16.6        17.0
    3       41.2       24.0        22.4
    4       73.7       17.5        17.8
    5       82.1       15.7        13.4
    6      100.7       14.7        12.5
    7       84.5       16.1        14.2
    8       61.7       16.0        14.3
    9      131.4       12.9        13.1
   10       56.3       14.6        16.0
   11       38.9       10.4        11.2
   12       34.5       18.7        20.4
   SP       90.6       19.9        20.2
   SU       92.6       18.7        17.2
   AU      101.3       17.1        19.6
   WI       32.1       18.3        16.6
    Y      118.4       18.1        18.4
 Critical value (significance level 0.05): 20.53
 Test statistics (TS) can be compared to the critical value.
```

REFERENCES

Mekis, E., Vincent, L., 2004: „New developments in the homogenization of precipitation and temperature series in Canada ", Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary; WMO, WCDMP-No. 56, pp. 39-62.

Szentimrey, T., 1999: „Multiple Analysis of Series for Homogenization (MASH)", Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2004: „Multiple Analysis of Series for Homogenization (MASH); Verification procedure for homogenized time series", Proceedings of the Fourth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary; WMO, WCDMP-No. 56, pp. 193-201.

Szentimrey, T., 2006: „Manual of homogenization software MASHv3.01", Hungarian Meteorological Service, p. 61.

Vincent, LA, Zhang, X, Bonsal, BR, Hogg, WD., 2002: „Homogenization of daily temperatures over Canada", *Journal of Climate* **15:** 1322-1334.

TESTING OF HOMOGENISATION METHODS PURPOSES, TOOLS, AND PROBLEMS OF IMPLEMENTATION

Peter Domonkos
H-1155 Budapest, Vasvári P. u. 5, Hungary
Tel: 36-209911774
dpeterfree@freemail.hu

1. INTRODUCTION

There has been a large number of homogenisation methods developed in the latest decades, for correcting observed climatological time series. However, only few papers analyse their practical efficiency with comparative tests. The earlier test procedures (*Easterling* and *Peterson*, 1995; *Ducré-Robitaille* et al., 2003; *Syrakova*, 2003; *Štěpánek*, 2004) were applied on relatively few homogenisation methods, and the degree of similarity of statistical properties between observed and simulated data sets were hardly or not at all examined.

In my opinion a high degree of similarity between the simulated and observed time series must be implemented, otherwise any test-procedure may bring misleading results. As the properties of observed climatological time series are quite diverse, and even the same homogenisation procedures may be applied with different parameterisations and supplementary parts (e.g. in relation to filtering outlier values or automatic consideration of metadata), it is not a simple task to construct a really usable comparative test procedure. In this paper some principles and initial steps are discussed, and some preliminary results are presented.

2. PRINCIPLES

A complete homogenisation procedure comprises several segments. The first step is usually the selection of set of time series which can be examined together. It is followed by filtering of outliers, the creation of reference series and calculation of relative time series, the use of statistical methods for finding inhomogeneities [= IH hereafter] in relative time series, etc. In this study the efficiency of one segment is examined only: what is the capability of the applied statistical method to find the timing of the IHs and assess the optimal corrections. From combined homogenisation procedures (e.g. MASH) segments of this type were separated before testing.

Further principles are as follows:

i) Only homogenisation methods having the following properties are tested: a) objective and reproducible, b) widely used in climatology c) its usable mathematical description is easily available in climatological journals or in the issues of previous homogenisation seminars. I constructed the necessary computer programs relying on these sources, with one exception (MASH, see later).

ii) Tests are fulfilled on large simulated data sets. The statistical characteristics of relative time series from observations are closely approached by those of simulated data sets through the thorough preparation of the simulation method.

iii) Methods, usually applied for finding real jumps in climate, are tested as well, since finding climatic jumps is a twin-task of homogenisation, both require the same type of statistical tools.

iv) In the present phase only relative time series with at least 0.4 autocorrelation are examined.

v) Segments of homogenisation procedures those comprise purely objective steps are tested. As in the present phase the way of creating reference series is not tested, the same way of creating relative time series is supposed for each homogenisation method.

3. HOMOGENISATION METHODS UNDER TESTING

Seventeen versions of 11 objective homogenisation methods are under examination. Two of the examined methods (the Multiple Linear Regression and the later version of the Standard Normal Homogeneity Test) are able to detect both sudden shifts and gradual changes (trends), but most of them are for detecting sudden shifts only. The list below comprises the tested homogenisation methods, it begins with the simplest ones which are followed by the more and more complex types:

a) t-test [tta] (Ducré-Robitaille et al., 2003)
b) t-test [ttb] (Kyselý and Domonkos, 2006)
c) Buishand-test [Buia] (maximum of the absolute values of accumulated anomalies, Buishand, 1982)
d) Buishand-test [Buib] (difference between maximum and minimum values of accumulated anomalies, Buishand, 1982)
e) Standard Normal Homogeneity Test for shifts only [SNHa] (Alexandersson, 1986)
f) Wilcoxon Rank Sum test [WRS] (Karl and Williams, 1987)
g) Multiple Linear Regression [MLR] (Vincent, 1998)
h) Bayesian test (Ducré-Robitaille et al., 2003) with serial correlation analysis (Sneyers, 1999) [Baya]
i) Bayesian test (Ducré-Robitaille et al., 2003) with penalised maximum likelihood method for calculating number of change-points (Caussinus and Lyazrhi, 1997; Mestre, 2004) [Bayb]
j) Pettitt-test [Pett] (Pettitt, 1979)
k) Mann-Kendall test [M-K] (Aesawy and Hasanean, 1998)
l) method of Mestre [Mest] (Mestre, 2004)
m) method of Mestre with parameterised minimum unit-length [Mesb]
n) Standard Normal Homogeneity Test for shifts and trends [SNHT] (Alexandersson and Moberg, 1997)
o) Easterling-Peterson test [East] (Easterling and Peterson, 1995)
p) Multiple Analysis of Series for Homogenisation [MASH] (Szentimrey, 1999)
q) Multiple Analysis of Series for Homogenisation with parameterised minimum unit-length [MASb]

In certain cases I made some supplements or modifications relative to the original homogenisation methods. Parameterised versions for Mestre-method and MASH are produced, because all the selected homogenisation methods are tested with various minimum unit-lengths. Minimum unit-length [denoted by u] means the shortest length of periods to which distinct statistical characteristics can be assigned during the search of IHs.

In most of the methods it is not fixed by original instructions, but it is definitely 1 single value in the original Mestre-method and in the original MASH. ("Number of values" in a series is referred as "number of years" hereafter, because it is the most frequent case in homogenisation procedures.)

At the Bayesian test I found the description of significance-test part to be non-reproducible (Ducré-Robitaille et al. 2003), therefore two other significance-tests are applied instead, producing two versions of the method. The Mann-Kendall test is supplied with the test of serial correlations, because otherwise the frequency of first-type errors seemed to be too high in white noise processes. In t-test the standard deviations are estimated from the whole series, instead of from 5-year long sections of the series.

T-test, Mest, East and MASH methods have inner instructions how to deal with multiple IHs, but most of the methods do not contain such unambiguous instructions. Moberg and Alexandersson (1997) give recommendations what can be an optimal way of detecting multiple IHs in a procedure (namely in SNHT) which includes step-by-step detection of individual IHs. The application of these proposals is implied in all of the methods that have no other relevant instructions.

In case of MASH, I have the opportunity to use the program with the assistance of the original constructor (Tamás Szentimrey, Hungarian Meteorological Service).

Further details about nine of the eleven homogenisation methods can be found in Domonkos (2006).

4. PREPROCESSING

Before the beginning of efficiency tests, some preparatory steps must be done. This section presents how the simulation method was developed, and deals with the problem of finding evaluation methods for measuring efficiencies of homogenisation methods.

4.1. Examination of an observed data set

One of the first tasks is to calculate statistical characteristics of an observed data set, since these characteristics must be compared with the same type of characteristics of simulated data sets before determining simulation method for producing data sets for the test procedure. Therefore, common statistical characteristics (moments, distribution, serial correlation), as well as mean characteristics of detected IHs are calculated for an observed data set. All the homogenisation methods are used, and the parameterisation of the individual methods is varied, in order to increase the number of the comparable characteristics. Before the beginning of the homogenisation procedures, relative time series are derived for each data series of the observed data set, according to the guides of Alexandersson and Moberg (1997).

The observed data set consists of 215 temperature and 112 precipitation time series. They are monthly or annual means (totals) of temperatures (precipitation), their length is 98 - 100 years, and most of them are originated from the observing network of the Hungarian Meteorological Service.

The obtained statistical characteristics varies according to the type of the examined meteorological element (temperature or precipitation), and the season of the year, but they can be well clustered in another way, according to the serial correlation of the relative time series. In the present work only time series with at least 0.4 serial correlations are examined, so mean statistical characteristics of 70 temperature and 2 precipitation (relative) series were selected for later comparisons with simulated characteristics.

Detailed description about the observed data set and the examinations performed on that is presented in Domonkos (2006).

4.2. Creation of the simulation method

An iteration technique was applied, namely simulation experiments were repeated many times in order to gain similar statistical characteristics of simulated data series to that of observed series. Though large numbers of statistical characteristics were compared, it appeared during the experiments that a set of statistical characteristics of detected IHs is partly more typical to the rate of factual small IHs relative to factual medium-size IHs, than to the factual mean frequency of all IHs. It is likely because of the fact that non of the homogenisation methods can detect very small IHs (see also Ducré-Robitaille et al., 2003; *Štěpánek*, 2004), thus direct information about the frequency of them cannot be obtained. To reduce the chance of possible biases because of an improper simulation method, the number of compared statistical characteristics was raised above 200. Yet it seemed that the function-type of the magnitude-distribution of small IHs can be chosen partly arbitrarily. Several results indicated that this distribution is close to that of the white noise, and eventually it was set to be just equal to the distribution of the white noise. Further properties of the simulation method were deduced from the results of the approximation experiments. The main findings are as follows: i) Frequency of small IHs is much higher, than that of IHs with medium-size or large magnitudes (see also Domonkos, 2006); ii) Frequency of short-term IHs is much higher, than IHs with long duration; iii) The examined time series can be well modelled composing the following elements: a) stationary white noise, b) change-points in certain time-points, c) gradual linear changes, d) platform-like changes (= pair of change-points with the same magnitudes, but with the opposite directions) and e) short-term platform-like changes. The latter is theoretically the same type as d), but I mention distinctly, because its frequency turned out to be surprisingly high, and its effect on the general characteristics of the time series is partly similar to that of the white noise. The length of the simulated series is always 100 years.
To determine a proper simulation method the forms of frequency distributions and suitable parameterisations were searched by iterations. The final simulation method is presented in Appendix I.

4.3. Handling of outliers

In the observed data set elements with higher than 4 standard deviations are very seldom. It is true both for the original and the relative time series. Anomalies higher than this threshold were corrected before the homogenisation procedure, so that they were substituted with the mean of the time series (i.e. with 0 anomalies). This practise is applied in all of the homogenisation procedures, and both for the observational and simulated time series.

4.4. Testing of the simulation method

Before the utilisation of the simulation method its properties were tested by the comparison of wide range of statistical characteristics of detected IHs. All the 17 homogenisation methods are involved. The compared characteristics are means of frequencies of detected change-points, means and standard deviations of the magnitudes of detected change-points, and in cases of MLR and SNHT the characteristics of detected gradual changes are also compared. Parameters of the applied minimum unit-length (u) and significance threshold (c) are varied, thus the number of the compared statistical characteristics is as high as 204.

Figure 1 presents the distribution of the differences between observed and simulated characteristics. Empty bars show the number of the differences belonging to the individual categories, while darkened bars show weighted sums of occurrences. The weights are the sample sizes of detected IHs, from which the characteristics of observed data set were calculated. It can be seen that vast majority of the differences are smaller than 10%, and they are never higher than 20%. The largest differences, which are higher than 10%, are partly explainable by the relatively low simple sizes belonging to these cases.

Distributions of serial correlations are also compared. The observational characteristics are calculated from the 72 relative time series whose serial correlations are higher than 0.4. Ten thousand time series were simulated, and the serial correlations were higher than 0.4 in 4533 cases. The comparison of relative frequencies of above 0.4 serial correlations are shown in fig. 2. The fitting is excellent; the distribution of the simulated serial correlations seems to be the smoothed version of the distribution of observational serial correlations.

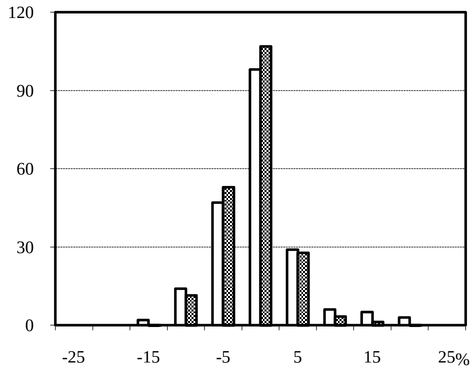


Fig. 1. Distribution of differences between the same type statistical characteristics of detected inhomogeneities for observed and simulated time series

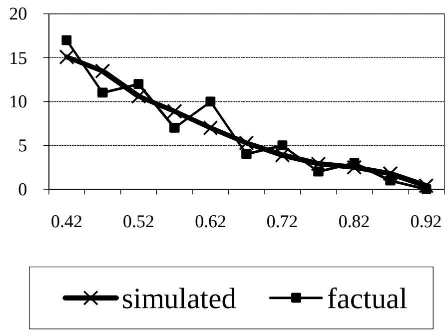


Fig. 2. Distributions of serial correlation of the examined time series

4.5. Measuring efficiency

The way of evaluating efficiency of a homogenisation method is not so simple as it could be expected over non-quantitative speculations. Let we examine a simple example: We have a relative time series with two change-points (and no more IHs, neither any noise, see fig. 3.a). The time-points and degrees of the shifts are i) year 40, +3 and ii) year 60, -2. In this idealised case all the values of the series would become 1 after a perfect correction. Method 'B' (fig. 3.b) detects precisely IH i), but fails to detect IH ii), thus it recommends a correction of +3 for the first 40 years of the series.

a) **b)** **c)**

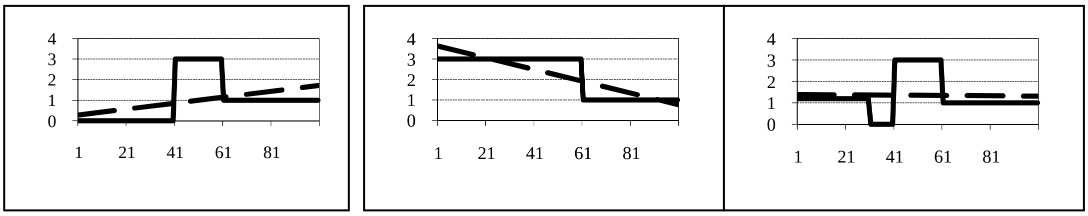


Fig. 3. The effect of IHs (a), and that of the adjustments of time series, relying on imperfect detections of IHs (b and c), on the detectable slope of linear change. (See more explanation in the text.)

Method ‘C’ (fig. 3.c) does not detect well the IHs, it finds IH neither around year 40, nor around year 60. However, method ‘C’ detects a change-point in year 30 with a shift of +1.2, and recommends a correction of +1.2 for the first 30 years of the series. The skill in detecting IHs is 50% for method ‘B’ and 0% for method ‘C’, but the reproduction of the trend of the series is good in case ‘C’ and false (worse, than without correction) in case ‘B’.

Two measures of efficiency are introduced: a) “Skill of change-point detection” evaluates the skill only in detecting relatively large shifts. (In certain cases it is the only purpose of the use of a homogenisation method.) b) “General efficiency” is a complex characteristic. It calculates the average of 14 elements characterising different kind of skills of the homogenisation methods. 4 elements are for evaluating the skill of detecting relatively large shifts, 4 elements measure the reliability of trends, 2 elements measure the reliability of the range of long-term changes, and 4 elements characterise the reliability of further properties of variability. The full description of the verification method is presented in Appendix II.

Efficiencies are expressed in percent unit, as the rate of the improvement relative to the perfect solution. The correction is perfect, if all the change-points and the gradual changes are precisely detected. If the distance of the corrected time series, relative to the perfect solution is just as large as that of the uncorrected series, the efficiency is zero. If the corrected series is worse than the uncorrected one, the efficiency is negative.

Some more specifications of the evaluation method are as follows: The perfect solutions are known only for simulated series. The calculation of perfect solutions takes into account the possibly previous modifications because of outlier values. However, there is a further factor complicating the calculation of efficiency. While all of the IHs of relative time series are usually considered to be the indications of errors in the candidate series, this rule may not be valid for very small IHs. Small IHs in relative time series can be caused by changes in climatic gradients or by the imperfectness of the reference series. Therefore a part of the small IHs is considered to be noise. Details are presented in Appendix III.

5. TESTING OF EFFICIENCIES OF HOMOGENISATION METHODS

The efficiency characteristics evaluated according to 4.5. were calculated for simulated data series, generated according to 4.2. All the 17 homogenisation methods and wide ranges of u and c parameters were applied. The number of generated series for each method and each of the applied parameterisations is always one thousand. The total number of experiments performed (considering all the homogenisation methods and applied parameterisations) with the simulation method of 4.2. is approximately $2 \cdot 10^7$. In addition, further experiments were made with certain modifications of the simulation method, in order to get an insight into the stability of the results. Results obtained with the base simulation method and with one modified version (few large IHs) are presented in this paper. The description of modified simulation method is included in Appendix I and III.

Efficiency values, obtained from adjacent values of c parameter applications, are smoothed with a Gaussian filter.

6. RESULTS

6.1. General efficiency

The Standard Normal Homogeneity Test is a very popular and the most often applied homogenisation method was in the last two decades. The earlier version (SNHa, for shifts only) was usually applied. Now, its efficiency can be checked. Fig. 4 shows the

distribution of general efficiency in the functions of minimum unit length (*u*) and significance threshold (*c*). *c* = 1 means the recommendation of Alexandersson and Moberg (1997) for 0.05 significance level, while *c* = 0 means the omission of applying significance threshold. The highest efficiency is 74.0%, and it belongs to the parameterisation of *u* = 6 and *c* = 0.59. We note that Moberg and Alexandersson (1997) search IHs in at least 10-year long series, thus the *u* in their application tends to be not shorter than 5 years. However, the optimal significance threshold is surprisingly low.
In fig. 4 the efficiency surface in a quite large area around the optimal parameterisation is flat, thus a wide range of parameter-pairs can be chosen maintaining near optimum efficiencies. However, application of a strict significance threshold or a pair of short *u* and very small *c* may result in a substantial decrease of efficiency.

Fig. 5 presents the distribution of efficiency for MASH (original MASH and MASb together) in the same way as fig. 4 does it for SNHa. MASH is a more sophisticated method; it has been worked out for detecting any possible combination of IHs in time series. The originally recommended parameterisation is *u* = 1 and *c* = 1 (Szentimrey, 1999). Results show that the optimum efficiency is slightly higher (75.6%), than that of SNHa, but the optimal parameterisation (*u* = 3 or 4, *c* = 0.75) differs from those recommended originally.

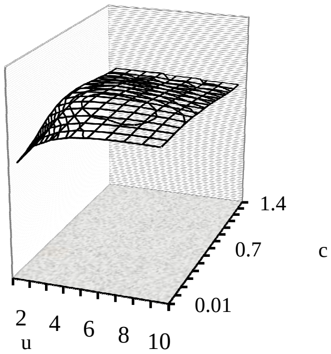


Fig. 4. Variation of general efficiency of SNHa with minimum unit-length (*u*) and significance threshold coefficient (*c*). Floor represents 0% and top of the wall means 100%.

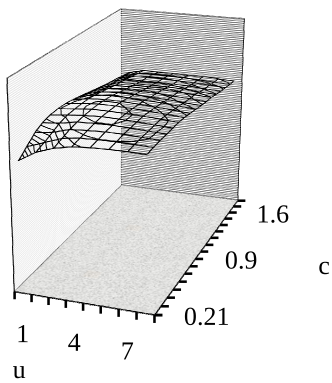


Fig. 5. Variation of general efficiency of MASH and MASb with minimum unit-length (*u*) and significance threshold coefficient (*c*).

As the area around the optimum flat again, pairs of parameters can freely be chosen from a rather wide range without an apparent loss of efficiency. Nevertheless, using the original parameterisation, the efficiency is 72.2% only.

Fig. 6 presents all the optimum efficiencies belonging to the examined methods. It can be seen that most of the values are within a narrow range (70 – 76%). According to these results the best methods are the Mesb, MASb, Mest, Bayb, MLR, MASH and SNHa, while the Mann-Kendall test and the t-test provide much lower general efficiency than the other methods.

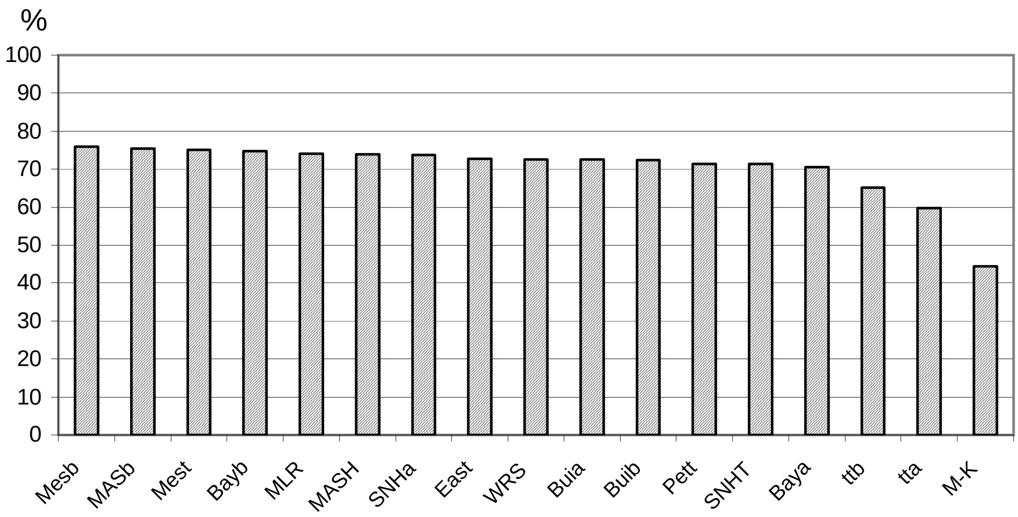


Fig. 6. General efficiencies of individual homogenisation methods, applying the optimal parameterisations.

The optimal significance thresholds are usually much lower, than the original recommendations, except for Mest and M-K. The omission of significance investigation provides the optimum efficiency in Baya, ttb, tta and in the first phase of East. The optimal *u* spreads over 3 - 10 years, and it tends to decrease with rising efficiency. Leaving out of consideration the very special case of M-K, the correlation between optimal minimum unit length and efficiency-optimum is –0.73.

6.2. Skill of change-points detection

Fig. 7 presents the order of the skill values of optimum parameterisations for all the examined methods, in the same way, as Fig. 6 shows the general efficiencies. All the values, except for that of M-K, range between 80 and 90%. Mesb and MASb are the most efficient two methods again, but the order substantially differs from the third place. tta seems to be very efficient for this task, while the rank of MLR is much weaker in this examination, than in the general efficiency results. The optimal parameterisation usually differs substantially from that of the general efficiencies, except for Mesb and MASb. The optimal unit length is 3 or 4 years for each of the methods. The optimal significance threshold is stricter in tta, ttb and East, but even lower, than for general efficiency for most of the methods. Omission of significance examination, or the use of a very low significance threshold is the optimal choice for Baya, Bayb, SNHa, SNHT, Buia, Buib, WRS, Pett and M-K.

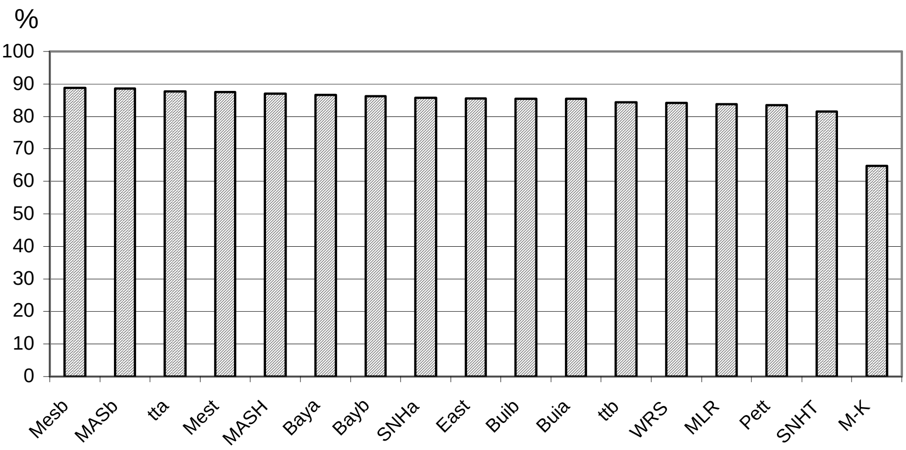


Fig. 7. Skill of change-point detection for individual homogenisation methods, applying the optimal parameterisations.

6.3. Case of few large inhomogeneities

Although the base simulation method imitates well the mean characteristics of the temperature time series observed in Hungary, the real time series characteristics are diverse. One of the frequent versions might be the "few large IHs" type, since in case of higher than average level of continuous quality control, the development of large shifts or biases may be prevented, while the features of small IH occurrences are likely the same.

Fig. 8 and 9 show the efficiency distributions of Mestre method (Mest and Mesb) for base dataset and few large IHs dataset, respectively. Comparing the results, it turns out that the efficiency much more depends on the characteristics of the candidate series and quality of the reference series, than on the applied homogenisation method. In case of Mesb the optimum efficiency drops from 76.0 to 53.8% owing to the change from base type to few large IHs type. The optimal *c* values are increased, but the shape of the distribution remained resembling otherwise.

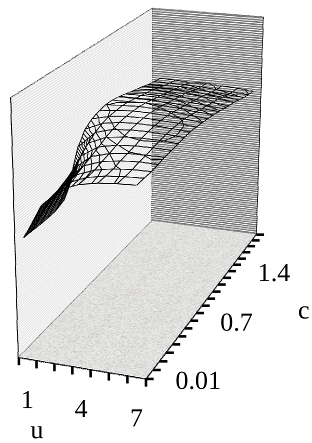


Fig. 8. Variation of general efficiency of Mest and Mesb with minimum unit-length (*u*) and significance threshold coefficient (*c*). Base-type data set.

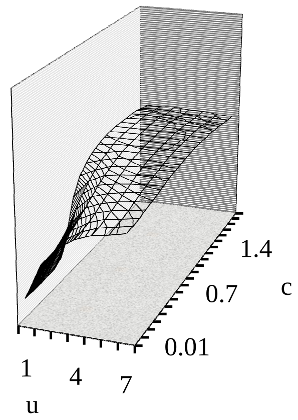


Fig. 9. Variation of general efficiency of Mest and Mesb with minimum unit-length (*u*) and significance threshold coefficient (*c*). Type of data set: few large IHs.

Fig. 10 presents another example of efficiency distribution for few large IHs type time series. Applying Buib method the optimal efficiency is slightly lower (51.2%), than that for the best methods. The distribution is conspicuously flat. The optimum significance threshold is near to the one for 0.1 probability of first type errors in white noise processes, but the choice of c = 0 (omission of significance test) results in efficiencies lower with few tenths of a percent only. The optimum u is 13 years, and the range of near optimal u values spreads over 7 to 18 years.

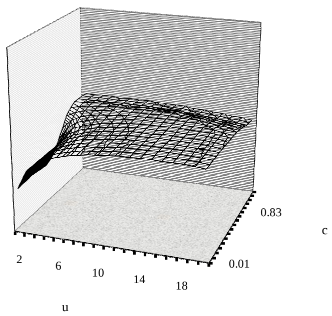


Fig. 10. Variation of general efficiency of Buib with minimum unit-length (*u*) and significance threshold coefficient (*c*). Type of data set: few large IHs.

In spite of the large changes in the measured efficiencies, the order of optimum efficiencies tends to be conservative against time series characteristics. Yet a little change can be noticed in the top places: for few large IHs type the highest optimum general efficiencies belong to 1) Mesb (53.8%), 2) Bayb (53.0%), 3) MASb (52.7%), while the order of the highest skills of change-point detection is 1) MASb (82.0%), 2) Mesb (81.7%) and 3) tta (79.6%).

7. DISCUSSION AND CONCLUSIONS

Large number of test experiments with simulated data set imitating relative time series from observations, with at least 0.4 serial correlation were performed for 17 homogenisation methods. All the calculated efficiencies belonging to realistic parameterisations are positive. It means that the application of any kind of homogenisation method generally improve the quality of climatological time series. (It does not mean, however, that substantial false corrections never happen.) Positive values of general efficiency calculated according to Appendix II. mean that the corrected time series are usually more appropriate for investigating climate variability, than the original series, independently from the kinds and sources of residual errors in the corrected series. In addition, most of the methods prove rather high efficiency in detecting and correcting change-points of relatively large magnitudes.

Most of the methods have similarly high optimum efficiency. It indicates that numerical values of efficiencies depend more on the properties of the data set under testing and on the chosen measure of efficiency, than on the homogenisation method. However, some exceptions have been found. Both types of the calculated efficiencies for the Mann-Kendall test, and the general efficiencies for the t-test are always substantially lower, than the efficiencies for the other methods.

It must be noted that the resemblance in efficiencies refers at first to the optimum values. Although the distribution of efficiencies is usually flat around the optimum efficiency, and thus parameter-pairs can be chosen from a rather large area, the use of original parameterisations recommended by the constructors of the methods sometimes results in considerably lower efficiencies. This problem is unusually serious for the Easterling-Peterson method. That method contains a two-phase examination of change-points with the application of a specific significance threshold for each phase. It has been turned out that the best choice is the omission of the significance-test in the first phase, and even a very weak significance threshold seems to be the best for the second phase (at least for time series with at least 0.4 serial correlations). The calculated general efficiencies for base type data set are 72.9% with the optimum parameterisation, but as low as 41.1% with the original parameterisation. Interestingly, the skill of change-point detection is high (84.1%) with the original parameterisation, which shows that the scale of reasonable expectations from a good homogenisation method must be wider than the reliable detection of large IHs.

The optimal parameterisation depends on the properties of the time series under examination. It is a problematic point, since the frequency and size-distribution of IHs are unknown in practice. However, stochastic-type information can be obtained by pre-examinations, for example serial correlation gives such information. A long-term purpose of efficiency-test investigations is to find methods and parameterisations whose application ensures high efficiencies for a wide range of time series properties.

The optimal parameterisation often contains very weak significance thresholds. It may have two explanations: a) The condition of minimum 0.4 serial correlation usually goes with a relatively high contamination of IHs, therefore the detected change-points can be approved without examining the statistical significance; b) Uncovered mid-size IHs might cause larger biases, than the application of unnecessary, but small corrections.

It seems that relatively complicated homogenisation methods are not always more efficient, than the simpler ones. For example, the earlier and simpler version of the SNHT (Alexandersson, 1986) produces higher efficiency, than its later development created for detecting both shifts and trends. The Easterling-Peterson method also seems to be an improper way of development, because the general efficiencies are low, except for parameterisations those are far from the original concept of the method.

During the evaluation of the results it must be considered that the category of relative time series with at least 0.4 serial correlations has been investigated so far, and the two, partly arbitrarily chosen measures of efficiency cannot be enough for drawing eventual conclusions.

Main conclusions of this paper are as follows:

Efficiency depends on the applied homogenisation method, but also on the preferred purpose(s) of the homogenisation.

Achievable efficiency is very sensitive to the statistical properties of the examined data set, but the rank-order of the optimum efficiencies among the different homogenisation methods tends to be conservative.

The efficiency testing procedure gives advice about the optimal values of significance threshold and minimum unit-length.

So far results show that for R1 > 0.4 type relative time series the Mestre method and the MASH are the most efficient homogenisation methods.

REFERENCES

Aesawy, A.M., Hasanean, H.M. 1998: Annual and seasonal climatic analysis of surface airtemperature variations at six southern Mediterranean stations. Theor. Appl. Climatol. 61,55-68.

Alexandersson, H. 1986: A homogeneity test applied to precipitation data. J. Climatol. 6, 661-675.

Alexandersson, H. and Moberg, A. 1997: Homogenization of Swedish temperature data. Part Homogeneity test for linear trends. Int. J. Climatol. 17, 25-34.

Buishand, T.A. 1982: Some methods for testing the homogeneity of rainfall records. J. Hydrology,58, 11-27.

Caussinus, H. and Lyazrhi, F. 1997: Choosing a linear model with a random number of change-points and outliers. Ann. Inst. Statist. Math., 49/4, 761-775.

Domonkos, P. 2006: Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation. Időjárás, 110, 63-87.

Ducré-Robitaille, J-F., Vincent, L.A. and Boulet, G. 2003: Comparison of techniques for detection of discontinuities in temperature series. Int. J. Climatol. 23, 1087-1101.

Easterling, D.R. and Peterson, T.C. 1995: A new method for detecting undocumented discontinuities in climatological time series. Int. J. Climatol. 15, 369-377.

Karl, T.R. and Williams, Jr. C.N. 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. J. Climate Appl. Meteor. 26, 1744-1763.

Kyselý, J. and Domonkos, P. 2006: Recent increase in persistence of atmospheric circulation over Europe: comparison with long-term variations since 1881. Int. J. Climatol. 26, 461-483.

Mestre, O. 2004: Detecting multiple change-points in a gaussian sample using dynamic programming. In: Fourth Seminar for Homogenization and Quality Control in Climatological Databases. WMO, WCDMP-No. 56, 89-92.

Moberg, A. and Alexandersson, H. 1997: Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. Int. J. Climatol. 17, 35-54.

Pettitt, A.N. 1979: A non-parametric approach to the change-point problem. Applied Statistics, 28, 126-135.

Sneyers, R. 1999: Homogenizing time series of climatological observations: the search and adjustment for inhomogeneities. Second Seminar for Homogenization of Surface Climatological Data. WMO, WCDMP-No. 41, 5-14.

Štěpánek, P. 2004: Homogenization of air temperature series in the Czech Republic during a period of instrumental measurements. In: Fourth Seminar for Homogenization and Quality Control in Climatological Databases. WMO, WCDMP-No. 56, 117-133.

Syrakova, M. 2003: Homogeneity analysis of climatological time series – experiments and problems. Időjárás, 107, 31-48.

Szentimrey, T. 1999: Multiple Analysis of Series for Homogenization (MASH). Second Seminar for Homogenization of Surface Climatological Data. WMO, WCDMP-No. 41, 27-46.
Vincent, L.A. 1998: A technique for the identification of inhomogeneities in Canadian temperature series. J. Climate, 11, 1094-1104.

APPENDIX I. Simulation of time series

a) Denotations

D – duration of the effect of IHs
G – random element of the standard Gaussian distribution
Int(..) integer part
K(..) function whose value is 0 or 1.
N – length of the time series (100 years)
p – parameter of time series, it characterises the rate of short-term IHs and the magnitude-distribution of IHs.
q – random element of the uniform distribution over the period [0,1)
r – parameter characterising serial correlation of series of IHs, its value is -0.5.
sign(..) function whose value is -1 or 1 according to the sign of the parent value
W – white noise process with standard Gaussian distribution $\mathbf{W}^{-1} = [w_1, w_2, ..., w_N]$
X – generated (target) time series: $\mathbf{X}^{-1} = [x_1, x_2, ..., x_N]$
Y – (time series of) accumulated effects of long-term IHs
Δ**Y** – (time series of) long-term IH events. Each event is paired with a specific year.
Z – (time series of) accumulated effects of short-term IHs
Δ**Z** – (time series of) short-term IH events. Each event is paired with a specific year.

Note: serial numbers of the elements of time series are indicated as indices of variables, but these indices are often omitted. Indices of *D*, *G*, K and *q* denote else, namely the repeated use of similar type variables/functions.

b) Simulation of data series of base data set

(i) 196-year long series are generated, and always the slices of years 48 - 147 are the target series.
(ii) IHs and noises are introduced in each year (but their values can be 0, naturally).
(iii) Types of the terms for introduction to time series: a) long-term IH (Δ*y*), b) short-term IH (Δ*z*) and c) white noise (*w*). Note: certain part of *y*- and *z*-type terms are handled as noise later (see Appendix III).
(iv) Forms of the IHs: a) sudden shift, b) gradual change, c) platform-like change, d) bias for one specific year. Form d) is a specific case of class c).

(v) Introduction of long-term IHs.
(v)/1: Size and direction of the IH
This term includes an IH whose magnitude can be large, with the probability given in K_1, as well as a small IH with the probability given in K_2:

$$\Delta y_i' = \mathrm{K}_1(q_1)\cdot \mathrm{sign}(0.5-q_2)\cdot(8+4p)\cdot q_3^{6+4p}) + \mathrm{K}_2(q_4)\cdot G_1, \quad (1)$$

where $\mathrm{K}_1(a) = 1$, if $a < 0.012$, and $\mathrm{K}_1(a) = 0$ otherwise; $\mathrm{K}_2(a) = 1$, if $a < 0.07$, and $\mathrm{K}_2(a) = 0$ otherwise; p has the same distribution as q does, but p is constant for a time series.

(v)/2: Form of the IH
The form of Δy_i' is (A) sudden shift, (B) gradual change or (C) platform-like change, with 0.4, 0.25 and 0.35 probability, respectively.

For (A)- and (B)-form IHs:

$$\Delta y_i = \sqrt{1-r^2}\cdot \Delta y_i' - r\cdot F \ , \qquad (2)$$

where $F = 0$ for the first (A)- or (B)-form IH of the series, and $F = \Delta y_k$ otherwise, k indicates the year of the previous introduction of (A)- or (B)-form IH.

For (C)-form IHs:

$$\Delta y_i = \Delta y_i' \qquad (3)$$

(v)/3: Calculation of the y_i components of the series
(A)-form IHs:

$$y_j = y_{j,-1} + \Delta y_i \text{ for each } j \in [i, N], \qquad (4)$$

where $y_{j,-1}$ denotes the value of term y_j before the actual adjustment.

For (B)- and (C)-form IHs duration-values must be paired at first. For B-form IHs

$$D_1 = 5 + 2\cdot \mathrm{Int}(48\cdot q_5^{1.5}), \qquad (5)$$

$$y_j = y_{j,-1} + \frac{(j-i+0.5D_1)\Delta y_i}{D_1} \text{ for each } j \in [i-0.5D_1, i+0.5D_1-1], \quad (6)$$

and for (C)-form IHs:

$$D_2 = \mathrm{Int}(30\cdot q_6^{1.5}), \qquad (7)$$

$$y_j = y_{j,-1} + \Delta y_i \text{ for each } j \in [i, i+D_2]. \qquad (8)$$

(vi) Introduction of short-term IHs
The size and the direction of this term is calculated by the same functions as those of long-term IHs (eq. 1), but the frequencies (determined by the K-functions) are different:

$$\Delta z_i' = \mathrm{K}_3(q_7)\cdot \mathrm{sign}(0.5-q_8)\cdot(8+4p)\cdot q_9^{6+4p}) + \mathrm{K}_4(q_{10})\cdot G_2, \quad (9)$$

where $\mathrm{K}_3(a)$=1, if $a < 0.04 - 0.03p$, and $\mathrm{K}_3(a) = 0$ otherwise; $\mathrm{K}_4(a)$=1, if $a < 0.5 - 0.4p$, and $\mathrm{K}_4(a) = 0$ otherwise.

$$\Delta z_i = \sqrt{1-r^2}\cdot \Delta z_i' - r\cdot z_{i,-1} \qquad (10)$$

The form of this term is always platform-like change.

$$D_3 = \mathrm{Int}\left(\frac{12 \cdot q_{11}^3}{1+0.3\left|\Delta z_i\right|}\right), \tag{11}$$

$$z_j = z_{j,-1} + \Delta z_i \quad \text{for each } \; j \in \left[i, i + D_3\right]. \tag{12}$$

(vii) Introduction of white noise term

$$w_i = G_3 \tag{13}$$

(viii) $\mathbf{X} = \mathbf{Y} + \mathbf{Z} + \mathbf{W}$ (14)

(ix) Serial correlation of **X** is calculated, and the series is added to the data set if the value is not lower than 0.4, but discarded otherwise.

c) Simulation of data series for the data set of few large IHs

The procedure is the same, as for the base data set, except for K_1 always equals to zero.

APPENDIX II. Calculation of efficiency

Time series are considered to be the sum of IH and noise components: **X** = -**V** + **W** (the series of positive *v* values indicates the theoretically perfect corrections). The series of estimated corrections by homogenisation methods is denoted with **U**.

a) Skill of change-point detection

It takes into account 6 characteristics. All of them have 2 values: good (1) or bad (0). The skill is calculated as the weighted average of the six characteristics. The weights are 1-1 for characteristics 1) – 4), and 2-2 for characteristics 5) and 6).

The way of evaluation is presented for negative shifts only (there is no logical difference in the evaluation of positive shifts).

1) Detection of large, long-term changes. The form of the IH:

$$\frac{1}{k}\left(\sum_{j=i}^{i+k-1} v_j - \sum_{j=i-k}^{i-1} v_j\right) \geq 3 \quad \text{for each } \; k \in \left[1, 2, ..., 10\right] \tag{15}$$

The best estimation of the change-point (denoted with *v**):

$$v_i^* = \frac{1}{3}\left(\sum_{j=i}^{i+2} v_j - \sum_{j=i-3}^{i-1} v_j\right) \tag{16}$$

The detection is good (1), if $\exists j,\, j \in \left[i-2, i+2\right]$ for which $v_i^* - 1 < u_j - u_{j-1} < v_i^* + 1$; and 0 otherwise.

2) Detection of large, short-term changes.

$$\frac{1}{k}\left(\sum_{j=i}^{i+k-1} v_j - \sum_{j=i-k}^{i-1} v_j\right) \geq 3 \quad \text{for each } k \in [1,2,3], \tag{18}$$

but $\exists m, m \in [4,5,6]$ for which $\frac{1}{m}\left(\sum_{j=i}^{i+m-1} v_j - \sum_{j=i-m}^{i-1} v_j\right) < 2.5$ (19)

The detection is good (1), if $\exists j, j \in [i-2, i+2]$, for which $u_j - u_{j-1} \geq 1$, and 0 otherwise.

3) Detection of medium-size, long-term changes.

$$\frac{1}{k}\left(\sum_{j=i}^{i+k-1} v_j - \sum_{j=i-k}^{i-1} v_j\right) \geq 1.5 \quad \text{for each } k \in [1,2,...,10] \tag{20}$$

The way of the verification is the same, as for case 1).

4) Detection of medium-size changes with very long duration.

$$\frac{1}{k}\left(\sum_{j=i}^{i+k-1} v_j - \sum_{j=i-k}^{i-1} v_j\right) \geq 1.5 \quad \text{for each } k \in [1,2,...,20] \tag{21}$$

The way of the verification is the same, as for case 1).

5) Change-point appears in the **U** series, and the detection is right.

$\Delta u_i = u_i - u_{i-1} \geq 1$; v_i^* of eq. (16) is positive;

$$\min(\Delta v) < \Delta u_i + 1, \tag{22}$$

$$\Delta u_i - 1 < \max(\Delta v), \tag{23}$$

$$\min(\Delta v) = \min(v_i, v_{i+1}, v_{i+2}) - \max(v_{i-1}, v_{i-2} v_{i-3}), \tag{24}$$

$$\max(\Delta v) = \max(v_i, v_{i+1}, v_{i+2}) - \min(v_{i-1}, v_{i-2} v_{i-3}), \tag{25}$$

6) Change-point appears in the **U** series, but the detection is false.

$\Delta u_i = u_i - u_{i-1} \geq 1$, but v_i^* of eq. (16) is not positive.

b) General efficiency

It takes into account 14 characteristics (d1, d2,…,d14). The verification results are expressed in percents of the perfect corrections, where the 0 means neither improvement, nor deterioration relative to the initial state. The general efficiency is the simple average of the 14 characteristics.

1) The same as the 1) of skill of change-point detection.

2) The same as the 3) of skill of change-point detection.

3) The same as the 5) of skill of change-point detection.

4) The same as the 6) of skill of change-point detection.

5) Difference in the slopes of the linear regressions (b) for the whole series.

$$d_5 = \left| b(u_{1-100}) - b(v_{1-100}) \right| \tag{26}$$

6) Difference in the slopes of the linear regressions for the last 50 years of the series.

$$d_6 = \left| b(u_{51-100}) - b(v_{51-100}) \right| \tag{27}$$

7) Rate of right decisions about the significance (at the 0.05 level) of the slopes of linear changes for the relative time series (X).

8) Rate of right decisions about the significance (at the 0.05 level) of the slopes of linear changes for the second half of the X series.

9) Difference in the ranges between the extremes of decade-averages.

$$d_9 = \frac{1}{10}\left(\max\left(\sum_{j=i}^{i+9} u_j \right) - \min\left(\sum_{j=k}^{k+9} u_j \right) \right) - \frac{1}{10}\left(\max\left(\sum_{j=l}^{l+9} v_j \right) - \min\left(\sum_{j=m}^{m+9} v_j \right) \right),$$

where $i,k,l,m \in [1,2,...N-9]$ (28)

10) Absolute value of the range-difference: $d_{10} = \left| d_9 \right|$ (29)

11) Combined characteristic for biases in size and/or in sign from the perfect corrections, and for time-lapses. First the combined characteristic for year i, (hi, i = [1,2,…N]) is performed. Let $u_i > v_i$ (the reverse case is handled with the same logic rules).

$$h_i = \min_k \left(g_k + u_i - \min\left(u_i, \max_j (v_j) \right) \right), \qquad (30)$$

where gk is a penalty-term of k-year lapse:

$$g_k = \exp(c_1(p \cdot k - c_2)) - c_3, \qquad (31)$$

c1 = 0.369, c2 = 3.297, c3 = 0.2962, p = 0.5; $j \in [j_1, j_2]$, $j_1 = \max(1, i-k)$,
$j_2 = \min(i+k, N)$ and k = [0,1,2,…,15].

$$d_{11} = \max_{i,j}(h_i - h_j), \quad i, j \in [1,2,..., N] \qquad (32)$$

12) The same as 11), but p = 1.

13) The same as 11), but p = 2.

14) SSE error-term: $d_{14} = \sum_{i=1}^{N} (u_i - v_i)^2$ (33)

APPENDIX III. NOISE-PART OF INHOMOGENEITIES

A part of long-term IHs (**Y**) and short-term IHs (**Z**) is not considered to be errors of the candidate series, so it is handled as noise. The rate of this type noise increases with decreasing IH-magnitudes, and it is higher for platform-like changes, than for lonely shifts and gradual changes. The probabilities of noise (*P*) for given IHs are determined according to the rules below:

1) Base data set, platform-like IHs.

$$P_1 = \max(0.6 - 0.4 \cdot |\Delta y_i|, 0), \qquad (34)$$

where Δy_i is determined by eq. (2) of App. I. Eq. (34) also applicable for Δ*z*-type terms.

2) Base data set, lonely shifts and gradual changes.

$$P_2 = \max(0.3 - 0.4 \cdot |\Delta y_i|, 0). \qquad (35)$$

3) Few large IHs, platform-like changes.

$$P_3 = \max(0.6 - 0.2 \cdot |\Delta y_i|, 0). \qquad (36)$$

4) Few large IHs, lonely shifts and gradual changes.

$$P_4 = \max\left(0.4 - \frac{4}{15} \cdot |\Delta y_i|, 0\right). \qquad (37)$$

HOMOGENIZATION OF AIR TEMPERATURE AND RELATIVE HUMIDITY MONTHLY MEANS OF INDIVIDUAL OBSERVATION HOURS IN THE AREA OF THE CZECH AND SLOVAK REPUBLIC

P. Štěpánek[1] , K. Mikulová[2]
[1] Czech Hydrometeorological Institute, Regional Office Brno, Czech Republic
[2] Slovak Hydrometeorological Institute, Slovak Republic,
petr.stepanek@chmi.cz, katarina.mikulova@shmu.sk

ABSTRACT

Homogenization of monthly averages of air temperature and relative humidity has been carried out for the area of the Czech and Slovak Republics for the period 1961-2005. Because of presence of a noise in the series, statistical homogeneity tests give their results with some portion of uncertainty. Using various statistical tests along with various types of reference series made it possible to considerably increase the number of homogeneity tests results for each tested series and thus to assess homogeneity more reliably. Homogenization was performed on individual hourly observations and comparison demonstrating the improvement of results compared to the homogenization of daily averages was made. Air temperature and relative humidity series were compared in order to help identify to what extent multi-element processing can help improve the homogenization of individual elements. All data processing and analysis were carried out using AnClim and ProcClimDB (softwares developed for automatic processing, analyzing, homogeneity testing and adjusting of climatological data).

INTRODUCTION

Long time series often suffer from non-climatic effects. It became well known and accepted fact that such inhomogeneous or erroneous series can lead to biased results in climatological time series analysis. Inhomogeneities may occur when stations are relocated and by changes of observer, instruments and observing procedures. This type of information should be documented in station metadata but there are numerous cases where such metadata is incomplete or missing, so we can rely then only upon statistical test results. A large array of statistical techniques has been developed to detect inhomogeneities in climatological time series. Various methods and different countries approaches are described e.g. in Peterson *et al.* (1998) and Szalai et al. (1999, 2004).

Different tests often identify different inhomogeneities (particularly for smaller amount of change), application of different types of reference series usually leads to different results as well, differences occur also among individual monthly, seasonal or annual series. In a number of instances, particularly where detections in series coincide were identified, adjustments can often be clearly justified, even in the absence of metadata. The existence of good quality reference series is very important for the detection of real inhomogeneities. That is why it is useful to include as many series for a particular area as possible, to help identify series abnormalities. In the case of air temperature, spatial correlations decrease with distance quite slowly, and so, it makes sense to analyze large areas. In this study the homogenization of both Czech and Slovak Republics series was considered useful and appropriate. To further increase the quality of homogenization, the number of test results was increased by testing monthly means of individual observation hours (i.e. those taken at 07:00, 14:00, 21:00 hours local time). Additionally, two meteorological elements, air temperature and relative humidity were mutually compared for detected inhomogeneities and were used when considering final adjustments of series. Single shift inhomogeneities are the most frequent in climatological time series and easiest to detect. Only single shift inhomogeneities were examined and adjusted for in this study so far. Data formatting and processing was performed using database software called ProcClimDB (Štěpánek 2006b). Homogeneity testing and time series analysis was conducted on AnClim software (Štěpánek, 2006a)

DATA CHARACTERIZATION

The Czech and Slovak Republics cover a total area of 128 km^2. Both Republics are mountainous. The Czech Republic ranges from 115 m to 1602 m at its highest peak (Sněžka). Despite being smaller, the Slovak Republic has a much greater height range, ranging from 94 m to 2655 m at Gerlachovský štít. From west to east the climatic influence of ocean diminishes and the continental influence progressively increases.

Air temperature and relative humidity were analyzed as series of monthly means of observations taken at 07:00, 14:00 and 21:00 hours local time, and daily averages. Stations with a minimum length of 25 years were selected. For the period since 1961, 230 stations measuring air temperature and 217 stations measuring relative humidity were available. The mean minimum distance between stations is 18.6 km, and means altitude 448 m, (median 380 m). Nine stations are situated above 1000 m a.s.l., and 4 above 1500 m a.s.l.

Because of processing of large number of stations, different methods of homogenization (e.g. different reference series, different tests) were examined in smaller areas such as Southern Moravia (Czech Republic) and Western Slovakia (Slovak Republic), see Fig. 1. From the results of these areas, the most useful types of reference series and homogeneity tests were selected and consecutively applied to the whole area of the Czech and Slovak Republics.

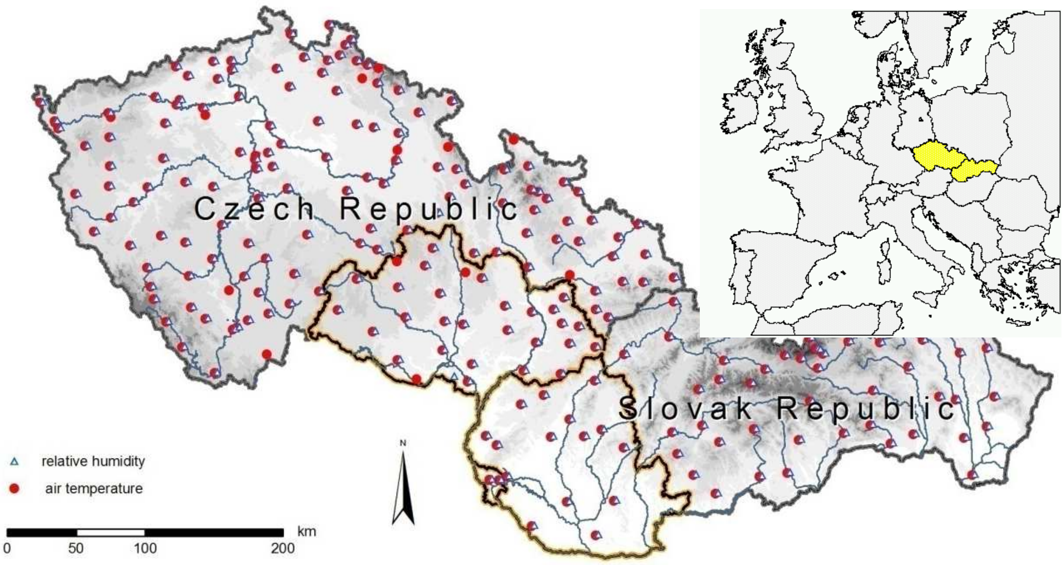


Fig. 1. Area of interest (Czech and Slovak Republics) with marked borders of the test area: Southern Moravia and Western Slovakia. Right top: location within Europe (at a different scale)

Air temperature was found to correlate very well throughout the Czech and Slovak Republics (see Fig. 2). Medians of correlation coefficients (from all the stations) vary only around 0.9 for all months in case of observation hour 14:00,

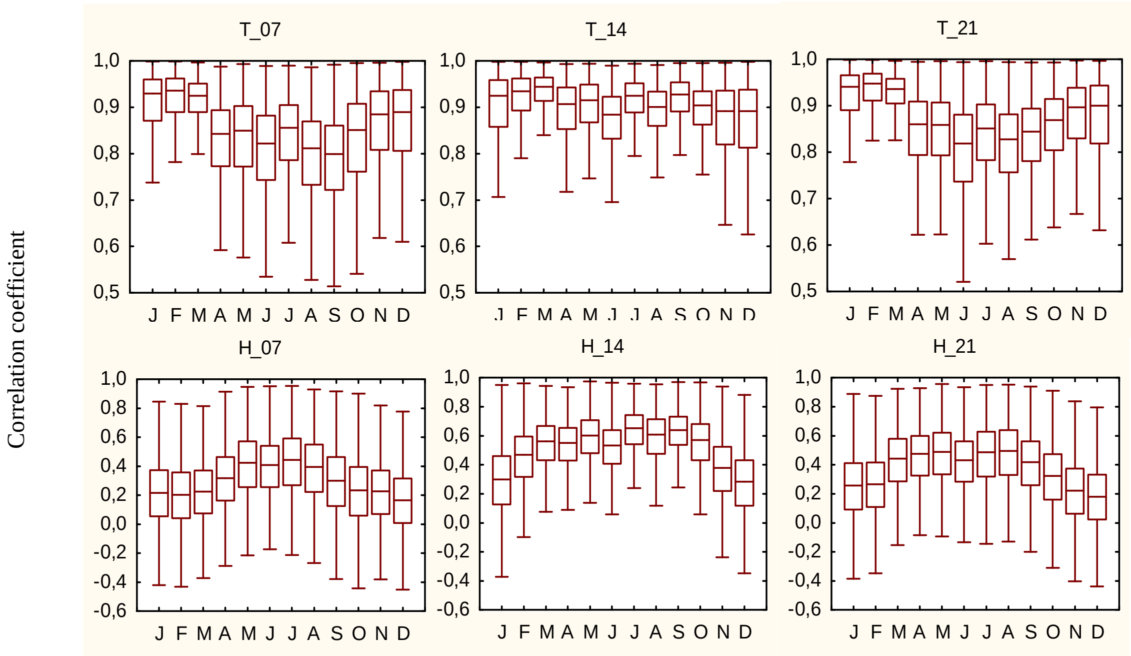


Fig. 2. Correlation coefficients for individual observation hours (07:00, 14:00, 21:00), for air temperature (T) and relative humidity (H) (using 25.420, resp. 22.595 station pairs - values)

and drop to 0.8 in case of hours 07:00 and 21:00. Correlations were lower in summer months and higher in winter. Values of correlation coefficients for daily averages are comparable with the hour 14:00, i.e. values of medians vary around 0.9, and during winter they are even higher.   Relative humidity correlates better in summer than in winter, again the best for the hour 14:00. Daily averages are comparable with the hour 14:00, and their values of correlation coefficient are similar or higher than in case of individual observation hours. Relative humidity correlations decrease relatively quickly with distance, but the stations network was sufficiently dense to create a well correlated reference series (see Fig. 8).

1. METHODOLOGY DESCRIPTION

In the case of series with missing or incomplete metadata, only statistical tests for homogeneity are relied upon to identify inhomogeneities. Unfortunately using solely the results of statistical tests during homogenization is problematic due to the fact that the detected year of inhomogeneity is often given with some error, or not identified. Štěpánek (2004) demonstrated that the determination of the correct year of inhomogeneity for air temperature, where the difference was less than 0.5°C, occurred in less than half of the cases. In the remained of cases, false years of inhomogeneity were given, or the years were not detected. According to this result, inhomogeneities less than 0.5°C are likely to be difficult to detect.

Because of this uncertainty in the result of homogeneity testing, it was attempted to increase the reliability of inhomogeneity determination through processing as many test results for each candidate series as possible. Series of individual observation hours were used, and several statistical tests for homogeneity were applied, various types of reference series were calculated for each candidate series, and monthly series as well as seasonal and annual averages of the series were tested. By combining all of these it was possible to considerably increase the number of test results for each tested series, thereby increasing the reliability of the homogenization process. Through the statistical processing of a large number of test results, it was possible to calculate the probability of each inhomogeneity of a given series (probability calculated as a portion of count of detected inhomogeneities - for each year, group of years or whole series - as an amount of all theoretically possible detections). One of the advantages of this approach is that a sum count of all detected inhomogeneities out of all the theoretically possible detections in the series can be used for assessing quality of measurements of a particular station as a whole.

Processing of the series during quality control and homogenization included the following steps: detection, verification, where necessary the correction of outliers (extreme values), creation of reference series (various ways), homogeneity testing (using 3 homogeneity tests), inhomogeneities (years) determination according to test results and metadata, adjustment of inhomogeneities and, only at the end, filling missing values. These steps are outlined in the Fig. 3. and are further discussed in the text.

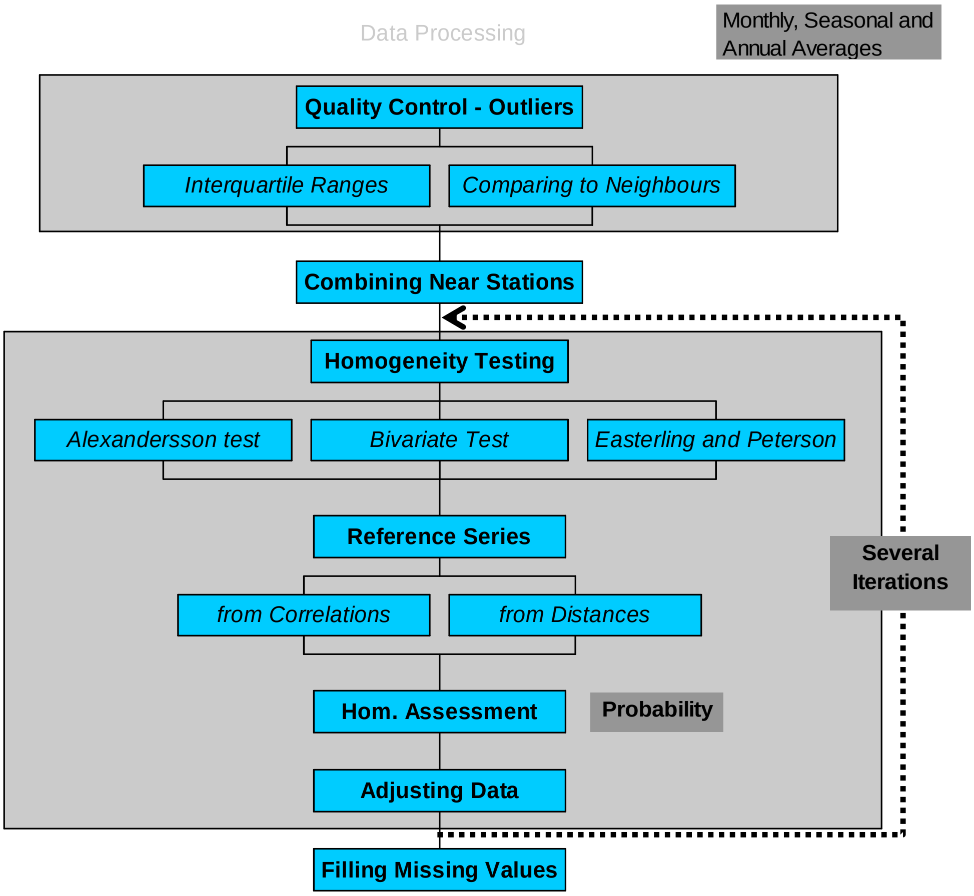


Fig. 3. Scheme of data processing during quality control and homogenization of the series

1.1. OUTLIERS IDENTIFICATION

Data quality control was carried out in two ways: by applying limits derived from interquartile ranges (it can be applied either to individual monthly series, or preferentially, by difference series between candidate and reference series), and, secondly, by comparing values to values of neighbouring stations

Where comparing neighbouring stations, the five best correlated neighbors were selected (correlations calculated form first difference series, see e.g. Peterson 1998), the values of correlation coefficients being at least 0.5, no limit for distance nor for altitude difference was applied. Only series with the same element and observation hour were selected. For outliers evaluation, the following characteristics were considered.

Counts of statistically significant different neighbours (compared to base station) exceeding confidence limit (0.95) were evaluated from difference series (neighbour and base station), the differences standardized to zero mean and standard deviation equal to one (to enable using standardized normal distribution), for each base station and month individually. Cases, where more than 75 % of neighbours significantly differed from the base station value, were visually checked. To help depict outliers, the values of neighbours were standardized with respect to base station average and standard deviation and also a new (theoretical) value for the base station was calculated - as weighted average from the standardized values of the neighbours (using 1/distances as weights, with power 1 which seems to be sufficient in case of air temperature, and 2 in case of relative humidity). Further, coefficient (multiply) of interquartile ranges (q75-q25) above q75 (or below q25) were evaluated (calculated from the standardized neighbours values), and applied to base station value. The reason for this was to assess the similarity of used neighbours values with regard to the outlier test value: the more values of neighbours are similar, the higher is the value of the coefficient.

The final decision on the removal of outliers was based on a percentage count of significantly different neighbours, the difference from the “expected value”, coefficient of interquartile range and finally value was conducted by visual (subjective) comparison of the standardized values of neighbours with the base station value.

1.2. COMBINING NEAR STATIONS MEASUREMENTS

In order to produce longer time-series, the neighboring station measurements (within 15 kilometers in case of temperature, and 10 kilometers in case of relative humidity, nearer stations having preference) was merged into one. A maximum gap of two combined series was allowed to be 4 years, the minimum length of reconstructed series was 25 years. In this instance, 14 stations were recorded and the year of merging then used as metadata information during series homogenization.

1.3. HOMOGENEITY TESTING

The AnClim software (Štěpánek, 2006a) was used to identify the inhomogeneities applying following tests for relative homogeneity (significance level α=0,05) on monthly, seasonal and annual data:

- Alexandersson test (SNHT for a single shift) (Alexandersson 1986, 1995)
- Bivariate test of Maronna and Yohai (Maronna and Yohai 1978, Potter 1981)
- Easterling and Peterson test (Easterling and Peterson, 1995)

To ensure that only one inhomogeneity was present in series when using Alexandersson or Bivariate test, a further modification was introduced into the AnClim software, which divides the series at the position of the found inhomogeneity and test the parts before and after the detected inhomogeneity separately. If no additional inhomogeneity was found in these two parts, we can rely upon the results of the given test for a whole length of the series (especially the significance of a test statistic).

These tests were applied for the whole studied area (the Czech and Slovak Republics). For the tested area, additional tests were also used to study differences in detection capabilities of individual tests and influence of various types of reference series (see chapter 4.2).

1.4. REFERENCE SERIES CALCULATION

In order to increase the number of homogeneity test results and thus better assess inhomogeneities in the series, two different calculations of reference series were performed:

- as an average from selected stations based on correlations
- as an average from selected stations based on distances

Each of these types of reference series has both advantages and disadvantages. By using correlations, the reference series created is the most similar to tested series (and thus suppressing variability in the differences/ratios series the best), but stations with similar inhomogeneities to the tested series can be selected. However this effect can be minimized by using first difference series for calculation of correlation coefficients, then inhomogeneities are manifested in one value (see e.g. Alexandersson, Moberg, 1996, Peterson 1998). For the latter type of reference series, by using distances, the geographical vicinity of the selected stations are preserved, but different climatic conditions even for near stations (due to different altitude etc.) can occur. Differences between reference series are further discussed e.g. by Mikulova and Stepanek (2004) or Stepanek (2005).

Weighted averages of neighbour series for reference series calculation were applied. The values of the selected neighbour stations were standardization to base station average and standard deviation to avoid problems with biased reference series. This can often happen in cases of missing data in of one of the neighbour series. No transformation of values (in case of air temperature and relative humidity) was applied to data.

In the first stage, a list of proposed neighbour stations was obtained. The list was subsequently checked, comparing correlation coefficients, distances, and also difference in stations altitude. This approved list was then finally used for the reference series calculation.

1.5. ASSESSMENT OF DETECTED INHOMOGENEITIES

The main criteria for determining the year of inhomogeneity was the probability of the given inhomogeneity, i.e. count of detections of a given year from all the testing of a given station expressed relatively to count of all theoretically possible detections. For detected inhomogeneities, a limit of 20% of all possible detections was used in cases where there was no information in metadata about the change. A limit of 10-15% was sufficient in cases where the inhomogeneity was in agreement with metadata. The count of detections for groups of years was also taken into account (some inhomogeneities started during the course of a year and thus manifested in 2 years at least). In cases where there was no mention in the metadata concerning a detected shift (which was most common), other sources information were used. Distribution of the given year within individual months or seasons, graphs of differences with reference series and some other characteristics, were all used for deciding whether the undocumented inhomogeneity could be regarded as "undoubtfully" proven and thus be corrected.

The mentioned decision limits were estimated subjectively, from selected set of stations, so that only clear inhomogeneities were corrected (an aim being not to “over-homogenize” the series). These limits are appropriate for the studied elements series of the analyzed area, for other elements or areas would have to be determined according to the given purposes.

1.6. ADJUSTMENT OF INHOMOGENEITIES

Adjustment of the detected inhomogeneities was carried out by means of reference series calculated as an average of five stations with the highest correlation coefficients to the series being adjusted (correlations were calculated again from the first difference series). The amount of change was estimated as a difference between averages calculated from difference series between the candidate and reference series from a period taken 20 years before and 20 years after the year being adjusted. The period was truncated in case another inhomogeneity within the period was encountered. These adjustments were applied to all monthly data. Where possible, the start of inhomogeneity was determined to a particular month.

Inhomogeneities within 4 years of the end of a series could not be adjusted. This happened relatively often in recent years, because of the transition to automatic measurements (being successively introduced since 1995). The parts of series with inhomogeneities near the ends of series had to be removed from further processing.

Various characteristics were analyzed before applying proposed adjustments including: increment of correlation coefficients between candidate and reference series after the adjustments, change of standard deviation in differences before and after the change, presence of linear trend, etc. In case of doubt, adjustments were not applied and the respective series was considered for removal from further processing.

Estimated adjustments are influenced by random errors in the series. To produce a smoother and physically more justifiable annual course of adjustments, weighted averages of the adjacent months were applied (with weights 1:2:1).

1.7. FURTHER CONSIDERATIONS

The above-mentioned steps (creating reference series, homogeneity testing, assessing and adjusting possible inhomogeneities) were performed in several iterations. In each iteration more precise results were obtained. The final adjustments of inhomogeneities were estimated from original data, taking into account inhomogeneities detected in all the previous iterations. It was necessary to use original data for the final correction (but used reference series were calculated from adjusted series in the last iteration), so that the final adjustments were estimated using periods without any inhomogeneities (a period taken for an adjustment was truncated when there was found another inhomogeneity in the series).

The filling of missing values was performed only after homogenization and adjustment of inhomogeneities in the series. The reason for this was to enable the new values to be estimated from data not influenced by possible shifts in the series. Moreover, when missing data are filled before homogenization, they may influence correct inhomogeneity detection (above all when a gap is longer than one year and there is an inhomogeneity - change of mean - near the position of the missing value). Filling the gaps was done by means of linear regression between filled value series (dependent variable) and a reference series (independent variable). Reference series were calculated as an average of five stations with the highest correlations with respect to the series with filled value. For the linear regression model, values 20 years before and 20 after the value being filled were used. Again, for assessing quality of the process, various statistics were monitored, e.g. differences of averages and standard deviations in periods before and after the gap.

A remaining question on the homogenization is the influence of the transition to automatic measurements which started in the station network of the Czech Republic in 1997 and in the Slovak Republic since 1995. For some of the series it is too early to assess the influence of automation, for other stations the effect was already very well detectable. The crucial problem is that the change do not manifest in the only data characteristic, such as arithmetic mean, but rather influencing several properties of the series, and moreover, these changes in series properties can not be usually linked linearly to the previous segment (before the change). That is why we need to possess long series (after the breaks) to be able to detect all the possible influences and, last but not least, to invent appropriate approaches for such analysis. Another difficulty is that there exist only few stations with comparative measurements, and stations from other sites are influenced by several other factors that are problematic to enumerate and thus to suppress.

2 SERIES FOR HOMOGENIZATION

2.1 Finding and removing outliers

For outlier identification, the approach described earlier in chapter 3.1 was applied. We tried to detect causes of anomalous monthly data by tracing the problems in daily data (the same method as mentioned above, applied to daily data within individual months), but due to huge database of processed values and shortage of time, we were not able to check all the detected values. This area will be the subject of further work. For the purposes of this study, approved errors were removed from further processing and were replaced by missing values. Considerably higher count of outliers occurs during summer months, in both processed elements (see Fig. 4)

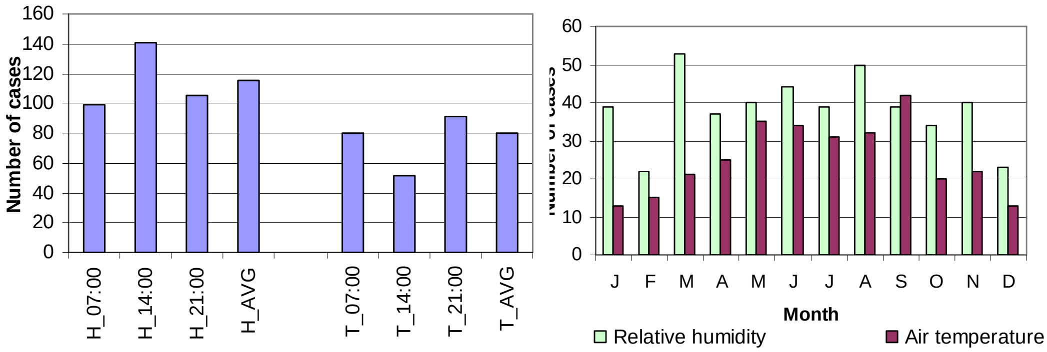


Fig. 4. ***Left*****: count of removed outliers for individual observation hours, for relative humidity (RV) and air temperature (T).** ***Right*****: count of removed outliers for individual months, for relative humidity and air temperature**

2.2 Verification of homogeneity detection procedure in the selected area

Due to the huge database (1.2 millions of monthly values for all the elements and observation hours within the whole area of the Czech and Slovak Republics, compared to "only" 240.000 months in the tested area), some approaches were tested in smaller areas, namely Southern Moravia and Western Slovakia (see Fig. 1).

We analyzed various types of reference series. These were created either by means of correlations or distances (using either simple or weighted mean), using either original series or series with removed outliers, using either the same element and observation hour of the neighbour stations as was that of candidate or using arbitrary observation hour from neighbours, or by using monthly or seasonal and annual averages. Altogether 18 different reference series were analyzed for each candidate. Reference series from distances and using arbitrary elements and observation hours of neighbours were not calculated because this option makes no sense.

Correlation coefficients for the different reference series are shown in Fig. 5. In case of air temperature, the medians of correlations are very similar, in winter there is no difference, in the summer months a maximum difference of five hundredths. The reference series for relative humidity differ, on the contrary, rather in the winter months, during the summer months the differences are again very small.

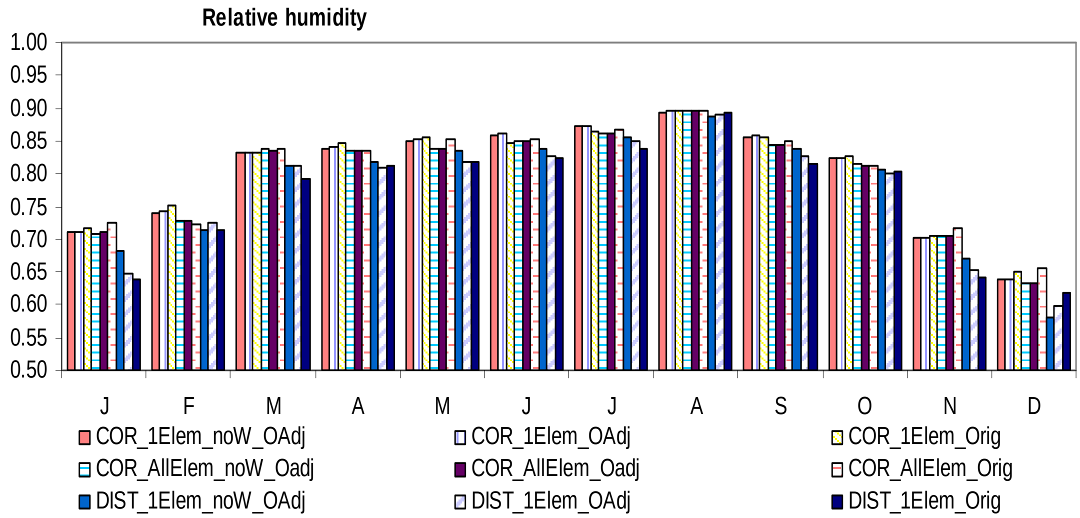


Fig. 5. Correlation coefficients (medians, calculated from 47 values) between various types of reference series and the tested series, for individual months. For explanation of used names of reference series see Fig. 6

Homogenization results for the questioned reference series and the Alexandersson test are shown in Fig. 6. The highest count of detected inhomogeneities occurs in the reference series created by means of distances, the same element and observation hour, using original series and unweighted mean. It is clear that this is caused by higher portion of random error in the tested candidate – reference difference series compared to the other types of reference series. Generally, series where outliers have not been removed give a higher count of detected inhomogeneities. The lowest count of detected inhomogeneities occurs in the reference series created by means of correlations. In Fig. 6 we can also see portion of the count of corrected to all detected inhomogeneities, which can be used for evaluation of the best reference series. The higher values of the portion are achieved by reference series created from correlations; the most efficient method to find true inhomogeneities is by using seasonal and annual averages for reference series calculation, series with removed outliers, and the same meteorological element and observation hour of the neighbour stations as in the tested series.

Following the presented results, for the homogenization of the series within the whole area of the Czech and Slovak Republics, reference series created both from correlations and distances, using neighbours with the same element and observation hour as that of candidate series and using weighted average were applied.

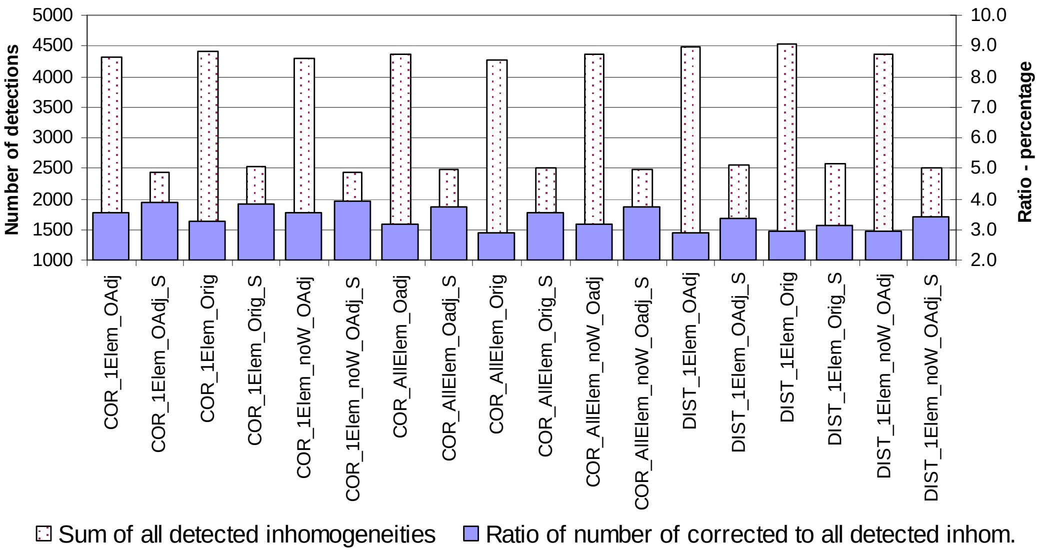


Fig. 6. Count of detected inhomogeneites for various types of reference series and ratio of the number of corrected inhomogeneities (100 adjusted elements and observation hours) to count of all detected inhomogeneities. *COR* – reference series created by means of correlations, *DIST* – by means of distances. *1Elem* – reference series created using neighbours with the same element and observation hour, *AllElem* – neighbours with different hours can be selected. *noW* – simple mean from neighbour series, otherwise weighted mean is used. *OAdj* – series with removed outliers, *Orig* – original series (no removal of detected outliers). *S* – seasonal and annual averages, otherwise monthly averages are used. Results for Alexandersson test

Various types of reference series give slightly different results due to random error present in the series. Where detections coincide, it is possible to better rely upon the test results. Including other types of reference series should lead only to a small improvement, therefore it seems more appropriate to gain further test results through testing individual monthly, seasonal and annual averages, testing individual observation hours etc.

2.3 Homogenization results for the whole studied area

As mentioned in chapter 3.1, we have used three tests for homogeneity, two types of reference series, monthly as well as seasonal and annual averages.

For the limits in which the values of correlation coefficients between candidate and reference series vary, for daily averages, see boxplots (i.e. median, lower and upper quartile and limits for outliers) in Fig. 7. Comparison of individual observation hours is shown in Fig. 8. For air temperature, the lowest values of medians (0.95) occur during summer months (for the hours 07:00 and 21:00), for the hour 14:00 they do not drop below 0.98, the same occurs for daily averages. In case of relative humidity, the correlation coefficients with reference series are markedly lower, mainly during winter months, but still usable for homogeneity testing. Again, values for the observation hour 14:00 are higher than those for 07:00 and 21:00, being comparable with daily averages.

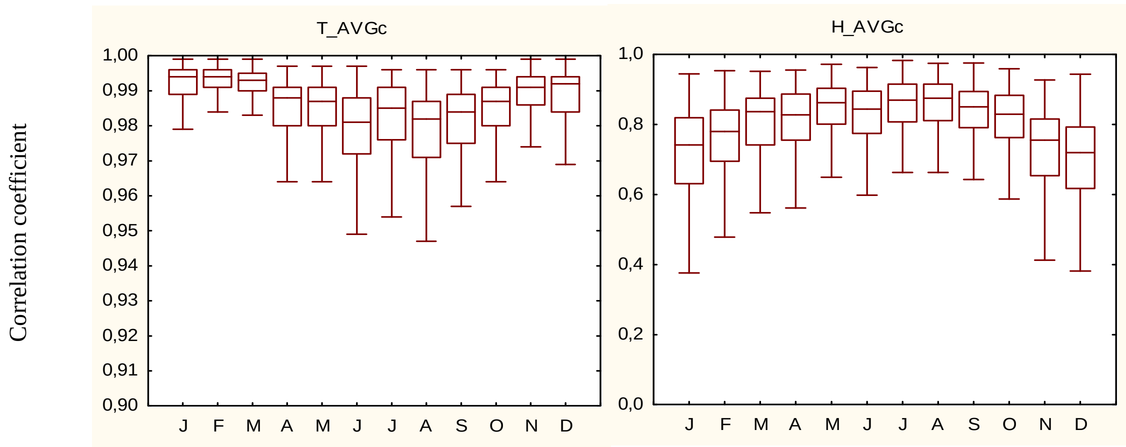


Fig. 7. Boxplots for correlation coefficients between candidate and reference series, applying correlations for reference series calculation, for air temperate (T, *Left*, from 227 values in each category) and relative humidity (H, *Right*, from 214 values)

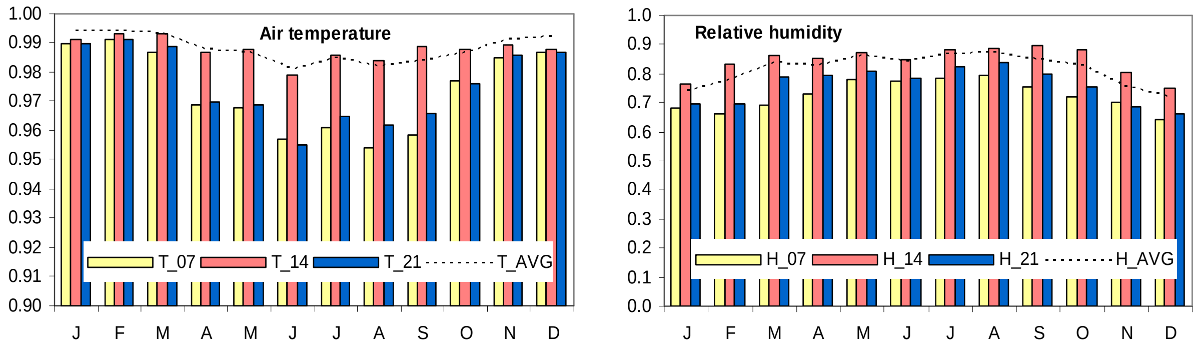


Fig. 8. Correlation coefficients between candidate and reference series (medians), for individual observation hours, applying correlations for reference series calculation (from 227, resp. 214 values for each category)

Fig. 9 shows the count of homogeneity tests detections for individual observation hours. Since data from the same place were used, both for air temperature and relative humidity, it is clear that inhomogeneities are more obviously detected in the air temperature series. One conclusion is that, effects such as station relocation manifest more profoundly in air temperature series (and within air temperature mainly during summer months). Nevertheless, an important role can also be played by lower correlations between candidate and reference series in case of relative humidity.

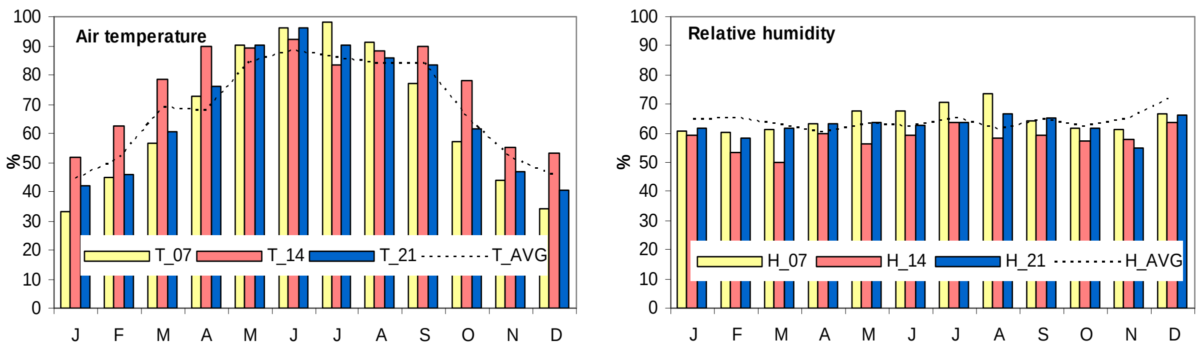


Fig. 9. Percentage of significant inhomogeneities (0.05) detected by used tests (SNHT test, Bivariate test, reference series created both by means of correlations and distances, altogether) related to the total number of series used. For individual months

In individual tests, the count of detections in individual years for Alexandersson, Bivariate and Easterling and Peterson tests is similar, if we do not consider ends of the tested series. The similarity of Easterling and Peterson test with Alexandersson one was a surprise because during simulations (for series with properties of air temperature series in the Czech Republic) that have been done previously (presented e.g. in Stepanek, 2005), the Easterling and Peterson test gave many more inhomogeneities, compared to Alexandersson test, that had no justification. This field is highlighted for further study, to better assess the relative power of these tests.

Counts of inhomogeneity detections – years – for various types of reference series show the same fluctuations, but for references series created by means of distance, the count of detections is about 10 percent lower for air temperature and 20 percent lower for relative humidity compared to reference series created by means of correlations.

Detected inhomogeneities were the subject of careful control before accepting final adjustments; this primarily took account of metadata, but also a number of auxiliary characteristics (see chapter 3.5). Fig. 10 shows final adjustments applied to the processed series. For air temperature, lower values of adjustments were applied for the hour 14:00 compared to the hours 07:00 and 21:00. The same behavior is valid for the improvement of correlation coefficients between candidate and reference series after realizing the adjustments. Relative humidity adjustments differ for individual observation hours and part of a year, e.g. in summer the highest values of adjustments occur for the hour 21:00, while in winter for the hour 14:00, the same course can be seen repeated for the correlations improvements after realizing adjustments.

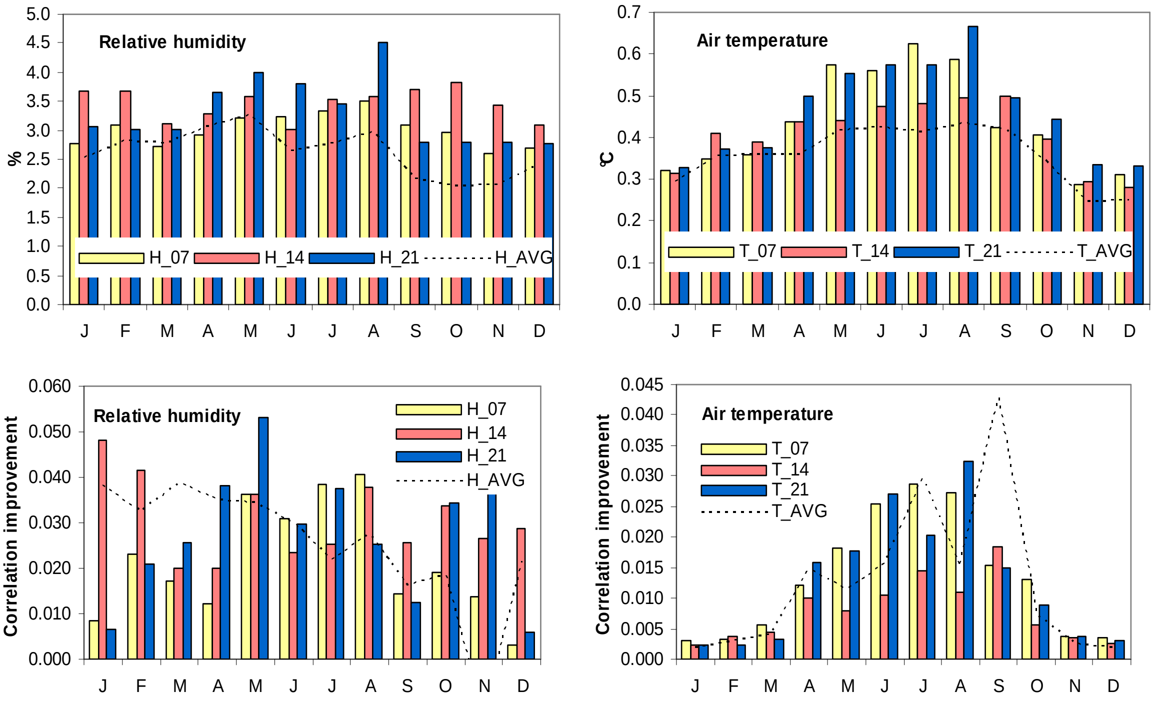


Fig. 10. Adjustments - averages of their absolute values and improvements of correlation coefficients (between tested and reference series) after realizing the adjustments, for air temperature and relative humidity, for individual months (using 40, resp. 32 values for each category)

In respect of air temperature (counts of detected inhomogeneities and amount of adjustments) the difference between summer and winter months can be explained by different influence of active surface upon the formation of air temperature regime in these distinct periods. In winter, prevailing circulation factors and reduced vegetation ensures the influence of effects leading to inhomogeneities (e.g. station relocation) is smaller, while in summer, resulting from prevailing radiation factors and increased volume of vegetation, the influence (of relocation for instance) is greater. The role of different active surfaces are also manifested in the fact that values of correlations are higher in winter months (climate conditions are similar for larger areas) in comparison with summer. Due to this characteristic, correlations were improved mainly in summer, in winter months the effect of homogenization was smaller (both for the adjustments and correlations improvement).
Relative humidity is a complex meteorological element influenced by many factors including air temperature, precipitation, wind speed, evaporation. This has the effect that explanation of inhomogeneity manifestation throughout a year is much difficult than in the case of air temperature. For instance, precipitation in the analyzed area is effected by station relocation primarily in winter, mainly due to the larger error in measurements connected with solid precipitation (manifested both in count of detected inhomogeneities and amount of adjustments).

Homogenization results obtained for air temperature can be generalized to a wider area (outside the analyzed area), since the spatial correlations of air temperature decrease with distance slowly. On the contrary, relative humidity correlations decreas rapidly with distance, so the presented results (annual course of inhomogeneities characteristics, etc.) can differ outside this study area.

3. SUMMARY

Our results indicate that analyzing series of individual observation hours improves the detectability of inhomogeneities. This is because inhomogeneities manifest in the different series in a different way: count of inhomogeneities, amount of change, correlations between reference and tested series (and thus delectability of inhomogeneities).

If we compare individual observation hours, inhomogeneities are better detected for the individual observation hour 14:00 – mainly because of higher correlations between candidate and reference series. Since inhomogeneities may manifest only in one of the observation hours and thus be masked in daily averages, performing homogenization on both individual observation hours and daily averages at the same time is recommended.

Valid for both the processed elements: station relocation and other effects that lead to inhomogeneities in the series are more profoundly manifested in summer months than in winter months in this study. In the case of air temperature, large differences of the used inhomogeneity characteristics occurred between individual months; the annual courses for relative humidity are, on the contrary, found to be much smoother. For relative humidity, the height of correlation coefficients (in summer) coincides well with the count of detected inhomogeneities, amount of adjustments, and correlation improvement (after adjustments). Inhomogeneity characteristics for air temperature were found to have a different annual course: higher correlations were found in winter but the count of detected inhomogeneities and amount of change were highest in summer.

Data processing and analysis was conducted using ProClimDB and AnClim software. This software is available from a server at http://www.klimahom.com/software/. Ongoing development of this software, e.g. connection with R software, is planned.

Acknowledgements

The authors would like to acknowledge Mr. Andy Moir, Brunel University, Uxbridge, for the undertaken English editing.

REFERENCES:

Alexandersson, A. (1986): A homogeneity test applied to precipitation data. J. Climatol., 6, 661-675.

Alexandersson, A. (1995): Homogenity Testing, Multiple Breaks and Trends. In: Proc. 6th Int. Meeting on Stat. Climatology, Galway, Ireland, 439-441.

Alexandersson, H., Moberg, A. (1996): Homogenization of Swedish Temperature Data. Part I: A Homogeneity Test for Linear Trends. in: Moberg, A: Temperature Variations in Sweden Since the 18th Century. The Department of Physical Geography, Stockholm University. Dissertation Series No. 5, 98 pp.

Easterling, D., R., Peterson, T., C. (1995): A new method for detecting undocumented discontinuities in climatological time series. Int. Journal of Climatology, vol. 15, 369-377.

Maronna, T., Yohai, V. J. (1978): A bivariate test for the detection of a systematic change in mean. J. Amer. Statist. Assoc., 73, 640-645.

Mikulová K, Štepánek P. (2004): Porovnanie referenčných radov pri testovaní homogenity relatívnej vlhkosti vzduchu vzhľadom na rôzne kritéria ich výberu, 7. konferencia mladých meteorológov a klimatológov, Zborník príspevkov, ISBN 80-88907-79-7, Bratislava

Peterson, T.C. (1998): Homogeneity adjustments of in situ atmospheric climate data: a review. Int. J. Climatol. 18, 1493-1517.

Potter, K.W. (1981): Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series. Mon. Wea. Rev., 109, 2040-2045.

Štěpánek, P. (2004): Homogenization of air temperature series in the Czech Republic during a period of instrumental measurements. In: Fourth seminar for homogenization and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Genova. 117-133.

Štěpánek, P. (2005): Variabilita teploty vzduchu na území České republiky v období přístrojových měření (Air Temperature Fluctuations in the Czech Republic in the Period of Instrumental Measurements). Dissertation, Geographical institute, Faculty of Sciences, Masaryk University, Brno. 136 pp.

Štěpánek, P. (2006a): ProClimDB – software for processing climatological datasets. CHMI, regional office Brno. http://www.klimahom.com/software/ProcData.html

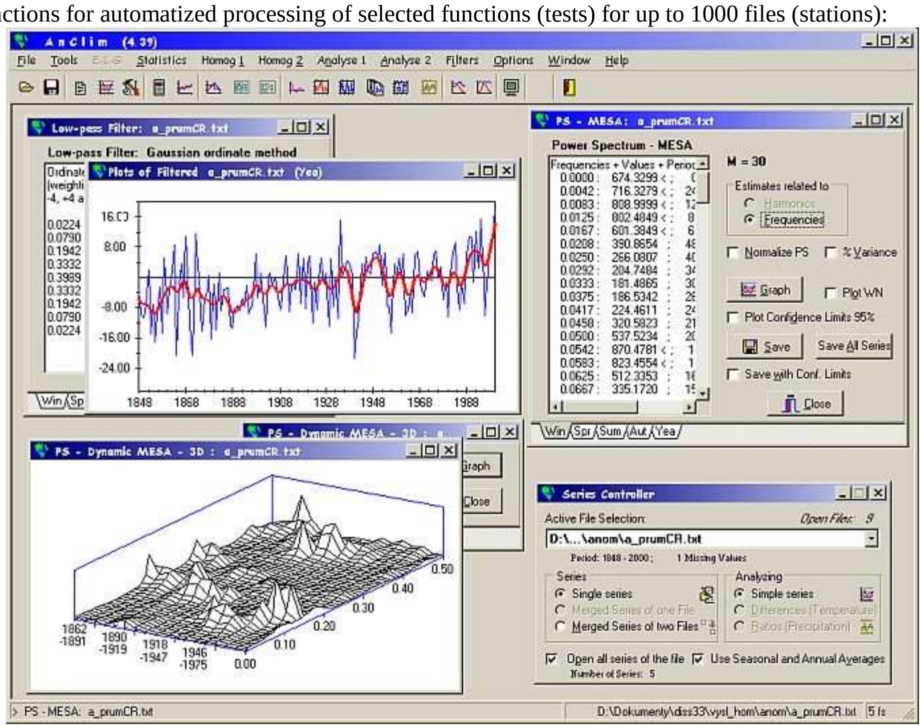
Štěpánek, P. (2006b): AnClim - software for time series analysis. Institute of Geography, Masaryk University, Brno, 1.47 MB
http://www.klimahom.com/software/AnClim.html

Szalai, S., Szentimrey, T. (2004): Proceedings of the Fourth seminar for homogenization and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Genova. 243 pp.

Szalai, S., Szentimrey, T., Szinell, C. (1999): Proceedings of the second seminar for homogenization of surface climatological data (Budapest, 9-13 November 1998). WCDMP-No. 41. WMO, Genova. 214 pp.

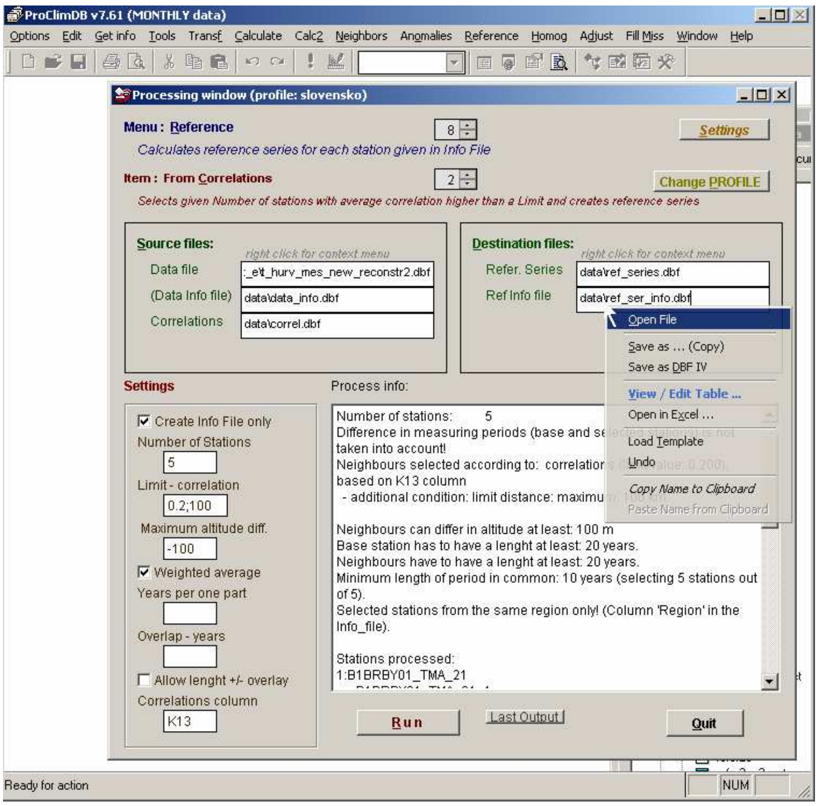
APPENDIX I. ANCLIM SOFTWARE DESCRIPTION

- **General characteristics**
  - o A comprehensive tool for processing monthly time series (from transformations through quality control and homogenization to time series analysis)
  - o Operates under Windows 95/98/NT/ME/2000/XP
  - o User friendly: a lot of graphical components, graphs clarifying the results, etc.
  - o Continuous development 1995-2006
- **Functionality**
  - o Series overview:
    - ▪ basic statistical characteristics, tests of randomness, outliers detection etc.
    - ▪ normal distribution testing, histograms
    - ▪ graphs of the series
  - o Regression models:
    - ▪ linear, polynomial regression.
    - ▪ multivariate linear regression graphs of the series
  - o Adjusting data:
    - ▪ replacing outliers, filling missing values
    - ▪ various transformations, converting series into anomalies from a mean, etc.
    - ▪ calculating differences/ratios of two series
    - ▪ switching between monthly or seasonal and annual averages
  - o Homogeneity testing:
    - ▪ absolute homogeneity tests
    - ▪ relative homogeneity tests (Alexandersson SNHT – various modifications, Bivariate test, Easterling and Peterson test, Vincent MLR, and others), creating reference series
    - ▪ adjustment of the inhomogeneities
  - o Time series analysis:
    - ▪ one series analysis (autocorrelations, power spectrum – MESA, dynamic MESA, etc.)
    - ▪ two series analysis (coherency analysis, etc.)
    - ▪ filtering the series (low-pass, band-pass, high-pass filters)
  - o Automation:
    - ▪ functions for automatized processing of selected functions (tests) for up to 1000 files (stations):



APPENDIX II. PROCLIMDB SOFTWARE DESCRIPTION

- **General characteristics**
  - Database software for processing climatological datasets (supports dbf IV files)
  - Two modes of processing: monthly or daily data
  - Automation of the processing (processing for a given list of stations, using all the stations in database)
  - Full control of the processing: many parameters for each option can be set, various outputs are created
  - Flexibility in modifying or adding new functions
- **Functionality**
  - Basic statistical characteristics computation, normal distribution testing, etc.
  - Finding outliers and extreme values
  - Neighbouring stations analysis (reconstructions, quality control, etc.)
  - Calculating correlation coefficients between all the pairs of a given list of stations
  - Reference series calculated as:
    - an average from the best correlated stations
    - an average from the nearest stations
    - an average of all stations available for a given year and month (regionally)
  - Processing output from the AnClim software homogeneity testing
  - Adjusting the series for inhomogeneities
  - Filling missing values:
    - from differences
    - using linear regression
  - Calculating monthly or seasonal and annual averages, calculating differences with a given reference series, etc.
  - Export to txt files, Excel, import from txt to dbf and other formats



A Quality Assurance System for Canadian Pressure Data

Hui Wan[1], Xiaolan L. Wang, and Val R. Swail
Climate Research Division
Atmospheric Science and Technology Directorate
Environment Canada, 4905 Dufferin Street
Toronto, ON, Canada M3H 5T4
E-mail: hui.wan@ec.gc.ca

(*Journal Applied Meteorology Climatology*, submitted)

ABSTRACT

In this study a comprehensive quality assurance (QA) system for both station and sea level pressure data is designed and applied to hourly pressure records (for 1953-2002) from 761 Canadian stations, to produce a high quality database of station and sea level pressures for various climate studies. The main principles of the QA system are described in detail, followed by a brief emphasis on the decision making and error correction algorithms. The general performance of the QA system and the main problems in the Canadian historical pressure database are discussed and illustrated through various examples.

The results show that there exist serious systematic errors (i.e., sudden changes in the mean, or mean-shifts) in the Canadian pressure database, which are caused either by the use of wrong station elevation values in the reduction of barometer readings to station or sea level pressure values (e.g., the "50 feet rule" or station relocation without updating the station elevation), or by transposition/swapping of station and sea level pressure values, or by mistakes made in the archive data ingestion or data recording/digitization processes (e.g., use of a wrong base number). Random errors also exist and are mainly due to transposition of two digits or miscoding of one or two digits. These errors must be corrected before the data are used in various climate studies, especially climate change related studies (e.g., assessment of climate trends and variability, weather regimes).

1. INTRODUCTION

Climate change has become an important issue, because increasing evidence suggests consistent warming trends over the past century, with a faster warming rate over land compared to oceans (Houghton et al. 2001). More and more efforts have been devoted to the assessment of climate change and their impacts. However, one needs long-term homogeneous records of climate data to characterize climate variability and climate change in the past, and to validate numerical model simulations. It is imperative to conduct quality assurance and homogenization of climate data before these data are used for various climate studies, especially climate change related studies.

Atmospheric circulation plays an essential role in the climate system because of its effects on the distribution of heat and moisture over the globe. Surface atmospheric pressure is an important variable that describes atmospheric circulation. Moreover, with the thermodynamic connection, the variations in surface pressure should reflect the variations in surface temperature. Therefore, analysis of surface atmospheric pressure is critical to our understanding of climate variability and climate change.

Several studies on the collection and analysis of atmospheric pressure data have been carried out lately. As a result, several good quality pressure data sets of global or regional coverage have been developed, mainly to provide vital inputs for numerical model studies of global climatic variations and changes (e.g., Smith and Reynolds, 2003; Kaplan et al., 2000; Allan et al., 1996, Trenberth and Paolino, 1980). Many data-quality related problems were found and corrected in these studies. These problems include data errors and discontinuities or inhomogeneities, and high-latitude station data problems (which are reportedly to have arisen from lack of data availability for the Arctic region).

In the mean time, there have been several studies using Canadian pressure data. Slonosky and Graham (2005) developed a Canadian monthly mean station pressure (SP) dataset with 71 stations that have data records for 50 to 130 years. They found strong correlations between the variability of atmosphere circulation and surface temperature anomalies. They also reported several major inhomogeneities in the dataset. Nkemdirim and Budikova (2001) examined trends in monthly mean SLP in western Canada using data from 51 stations for the period from 1956 to 1993, and reported a significant decline in annual mean and winter mean SLP over the Arctic.

However, the original records of surface atmospheric pressure are hourly measurements, from which the commonly used monthly or daily mean pressure values are derived. Unfortunately, the hourly pressure data archived in Environment Canada have not undergone a quality control (QC) or quality assurance (QA) procedure (only recorded with missing flags). Slonosky and Graham (2005) corrected some problems in their analysis of monthly pressure data, while Nkemdirim and Budikova (2001) did not (and hence their results are most likely unreliable). In order to produce a high quality, homogeneous pressure database for various climate studies, a comprehensive quality assurance and homogenization of Canadian pressure data is long overdue.

The necessity of applying a QA procedure to meteorological data has long been recognized. The earliest QA systems were developed, aiming at radiosonde data

(Gandin 1988; Collins and Gandin, 1990). However, more and more efforts have been put towards developing QA systems for high temporal resolution surface meteorological data, such as daily or hourly data (Kunkel et al. 1998, Graybeal et al. 2004, Shafer et al. 2000). A complex QA procedure consists of a number of checks on data; it uses the checking results comprehensively to determine whether or not a value is suspicious and how to correct the suspicious value if possible. Since not all flagged data are erroneous, a complex QA procedure should check all flagged data to screen out those most suspicious values (for correction or exclusion) and to remove flags from data that are deemed consistent with other reliable data. A modern complex QA system is not only to identify, but also to correct suspicious data whenever possible.

The Environment Canada (EC) digital archive contains pressure data from 1953 to date. For the early decades, data were digitized from original paper forms, without any quality control performed after digitization. Actually, even for the real time data (those from electronic reports), the QC procedure is quite limited according to the EC National Archive hourly data quality control documents published on the EC's website (Environment Canada 2004). Thus, a QA procedure for hourly pressure data is developed in this study with the goal of combining existing techniques and fitting them to Canadian historical data.

In this study, we develop a QA procedure for Canadian hourly pressure data. The data and the QA procedure are described in sections 2 and 3, respectively. Section 4 describes the decision making method, and section 5, the error correction algorithms. The corrected data series are analyzed in section 6, and this paper is completed with some concluding remarks in section 7.

2. DATA

Surface atmospheric pressure is usually recorded for both the station and the mean sea levels. Generally, the atmospheric pressure values at the station elevation are called station pressure (SP) and are calculated from the station barometer readings. Then, the mean sea level pressure (SLP) is derived from the SP, so that the barometric pressures at stations of different elevations can be compared at a common level (mean sea level) for synoptic purposes. Generally, SP data should be more reliable than SLP as fewer calculations are involved. However, SLP data have been used quite often for various purposes such as constructing atmospheric circulation indicators (e.g., Wright 1984, Jones et al. 1999), developing long-range climate forecast models (e.g., Christensen and Eilbert 1985), and analyzing severe weather phenomena (e.g., Wang et al. 2006, Alexander et al. 2005). Therefore, high quality data for both station and mean sea levels are needed for various studies.

Considering a further interest in producing a gridded pressure dataset, we should apply the QA system to as many stations as possible. There are 1085 stations available for both SP and SLP data in the EC data archive. Only stations with continuous records for at least one year and days with at least eight records were included in this study (although at most stations atmospheric pressure are reported hourly, with 24 measurements per day, some stations have only one report every 3- or 6-hour or have hourly reports for only part of day, e.g., from 03:00 to 16:00. The number of hours of pressure reports per day could vary from station to station and/or from one period to another). Because SLP data are derived from SP data, and we will use both elements for QA, the checking procedure will be applied to data only when both SP and SLP data are available. A total of 761 stations (see Fig. 1) are analyzed in the study.

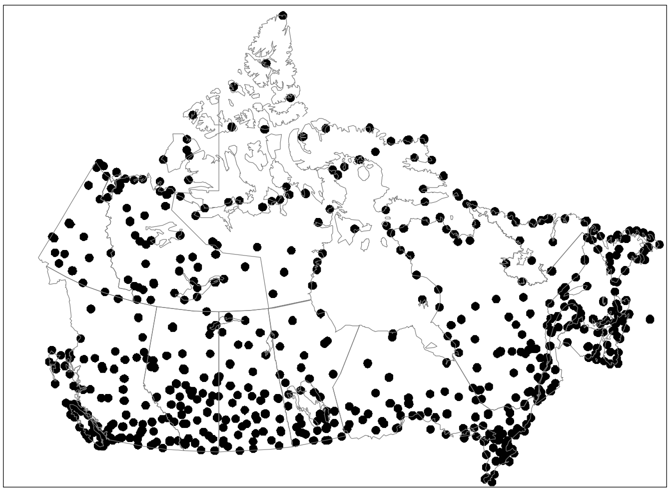


Figure 1. Location of stations analyzed in the study.

3. The QUALITY ASSURANCE SYSTEM

The QA system consists of five components. These include checking for upper and lower climatological thresholds/limits, for temporal pressure changes, and for hydrostatic, temporal, and internal consistencies. For each station, all valid (non-missing) values are subject to these five checks. Based on the results of these checks, a decision regarding acceptance or correction or rejection of the data is made.

a. Limits check (LC)

The climatological thresholds/limits check is a very commonly used checking procedure to wipe out outliers (e.g., Hubbard et al. 2005; Graybeal et al. 2004; Shafter et al. 2000). In this study, the climatological thresholds were determined as the lowest and highest values in the 1971-2000 period (for each station with at least 15 years of data in this period) that are associated with acceptable values of one-, two-, and three-hour pressure changes (i.e., those that are below the corresponding limits of pressure change; see section b below). If the lowest or highest hourly value is associated with unacceptable values of pressure change, we exclude it and seek for the second lowest or highest hourly value. This procedure is repeated until the acceptable climatological thresholds are found.

There are only 216 Canadian stations of at least 15 years of hourly pressure data in the 1971- 2000 period. Among these stations, the lower limits range from 940.2 hPa to 981.8 hPa for SLP, and from 846.7 hPa to 972.9 hPa for SP; while the upper limits range from 1040.2 hPa to 1079.5 hPa for SLP, and from 917.1 hPa to 1060.6 hPa for SP. Since the climatological limits were determined using 30 years of data, an arbitrary tolerance of 3.4 hPa (0.10 in. of Hg.) was added to the thresholds for each station. For a station of shorter data record, we use the lowest lower limit among its four "nearest" surrounding stations as its lower limit, and the highest upper limit, as its upper limit. Note that station elevation is also a factor we considered here: each of the four "nearest" stations must have an elevation difference from the short term station that is less than 200 m (otherwise it is replaced by the next nearest station; the 200 m limit is reasonable because it is only for choosing the climatological limits). This limit for difference in elevation is important for setting the climatological limits of station pressure, especially for elevated stations.

b. Pressure changes check (PC)

The limits for one-, two- and three-hour pressure changes taken from the EC hourly data quality control document (Environment Canada 2004) are used in this study. They are 3.9 hPa/hr, 6.9 hPa/2-hr, and 9.9 hPa/3-hr, respectively. These limits were developed in the early-mid 1990's by experienced meteorological technicians. Note that, for very rare events, the true pressure tendency could exceed these limits [e.g., Le Blancq (2003) reported that 3-hourly station pressure tendency was 28.9 hPa on 11 February 2003 from 10:00 to 13:00 at Sable Island, Nova Scotia, Canada]. A flag was issued to a datum if at least one of the associated pressure changes exceeds its limit.

c. Internal consistency check (IC)

The physical relationship that $SP < SLP$ is evaluated for stations with non-zero station elevation. Both *SP* and *SLP* data are flagged if this relationship is broken down. However, consecutive flags due to an identical value of *SP* and *SLP* are a special issue here. Slonosky and Graham (2005) reported discontinuities in *SP* data series that are due to a change in the definition of "station elevation". The latest edition of the “Manual of Surface Weather Observation” (Environment Canada 1977) states that “prior to 1 January 1977 the term ‘established elevation’ was used” and that “an established elevation of zero metres (MSL) was assigned to all stations where the cistern elevation was less than 15 metres (50 feet)”. As a consequence, the station pressure and the sea level pressure were identical at these stations before January 1977. Therefore, a special flag is activated when identical values of *SP* and *SLP* are found for at least one month. Actually, this “50 feet rule” problem could also lead us to flag a long run of consecutive hourly records during the hydrostatic check described below.

d. Hydrostatic check (HC)

The hydrostatic check has been used routinely in upper-air radiosonde data quality control (Gandin 1988, Collins and Gandin 1990). It plays a crucial role in identifying errors of height or pressure or temperature at mandatory isobaric surfaces. We use it here to detect errors in both station and mean sea level pressure data.

For station pressure P_z and sea level pressure P_o, the hydrostatic check is based on the hydrostatic model:

$$Z = \ln\frac{P_0}{P_z} \times (T_0 + \overline{T}_{dry}) \Big/ \left(\frac{g}{R} - \frac{a}{2}\ln\frac{P_0}{P_z}\right) \qquad (1)$$

where Z is the station elevation (in meters), R is the gas constant for dry air, T_0 =273.15*K*, g is the acceleration of gravity, a is the standard lapse rate (0.0065℃/meter), and $\overline{T}_{dry}$ is the average of the current dry bulb temperature and the dry bulb temperature recorded 12-hour earlier (unit: ℃).

However, since November 1976, the following formula is used in Canada for calculation of mean sea level pressure (Savdie 1982; WMO 1954):

$$P_0 = P_z \exp\left(\frac{gZ}{RT_{mv}}\right) \qquad (2a)$$

$$T_{mv} = (T_0 + \overline{T}_{dry}) + \frac{aZ}{2} + e_s C_h(Z) + F(\overline{T}_{dry}) \qquad (2b)$$

where

$$e_s = (\overline{T}_{dry} + T_0)^{-0.00014\overline{T}_{dry}^{\,2} + 0.0116\overline{T}_{dry} + 0.279}$$

$$C_h(Z) = 2.8322\times10^{-9} Z^2 + 2.225\times10^{-5} Z + 0.10743$$

$$F(\overline{T}_{dry}) = b_1\overline{T}_{dry}^{\,2} + b_2\overline{T}_{dry} + b_3$$

The third term in (2b) represents a humidity correction, where e_s is the surface vapor pressure and $C_h(Z)$ is the humidity correction factor (a function of Z). The last term accounts for correction of plateau effects, and b_1, b_2, b_3 are plateau correction parameters specific to each station. Note that neglecting humidity and plateau correction, the combination of equations (2a) and (2b) is equivalent to (1).

The hydrostatic residuals R_z are defined as

$$R_z = Z_m - Z \,,$$

(3)

where Z is the recorded station elevation (taken from station history metadata and hence can reasonably be deemed correct), and Z_m is the estimation of the station elevation by substituting the related hourly P_o or P_z or $\overline{T}_{dry}$ values in model (1). In the absence of data error(s), R_z values shall be very close to zero. However, a tolerance of R_z is used here to allow for small errors in the value of P_o or P_z, or even in the dry bulb temperature $\overline{T}_{dry}$ or the recorded station elevation Z (but undocumented large elevation changes can still be identified by this check). In this study, we assume that the recorded hourly dry bulb temperature values are correct. All hourly pressure data (both P_o and P_z) associated with a R_z value that is greater than its tolerance are flagged as a result of this hydrostatic check.

For each station, the tolerance of R_z is determined by

$$\mu - \gamma\sigma \le R_z \le \mu + \gamma\sigma \tag{4}$$

where μ and σ are the mean and standard deviation of the hourly R_z time series, respectively. A value passes the hydrostatic check if the above relationship holds. The value of γ is determined by analyzing the time series of R_z for each station, separately. Generally, in the absence of errors, both μ and σ should be near zero (cf. Fig. 2a). However, it was found that many data problems and errors could influence the estimates of μ and σ. Therefore, it is impossible to use one single value of γ for all stations. For example, as shown in Fig. 2b, a clear step was found in the time series of residuals R_z on 3 October 1965, with most of the residuals (in absolute value) before 1965 equal to the station elevation (19.2m), which is apparently an error caused by the "50 feet rule" problem (cf. subsection 3c). An improper value of γ could lead to flag all values of both P_z and P_o before October 1965.

We also encountered a number of cases in which the σ values are large (cf. Fig. 2c). Further investigation reveals that all these cases are associated with highly elevated stations (i.e., stations of very high elevation; e.g., the elevation of Old Glory Mountain is 2347m), which indicates that this very likely reflects the problem with the sea level pressure reduction (cf. Mohr 2004 and Pauley 1998): Reduction of station pressure to the mean sea level assumes a fictitious air column between the height of the station and the mean sea level. The air temperature decreases with increasing height from the surface. However, the mean temperature of the fictitious air column is unknown, and is usually approximated in Canada by using a standard temperature lapse rate and the dry bulb temperatures recorded now and 12 hours earlier (cf. Savdie 1982). Also, a plateau correction has been added since November 1976 for all stations in Canada (Savdie 1982), aiming to get approximately the same amplitude of the annual variation of sea level pressure at all stations, regardless of their elevation (WMO 1964; Mohr 2004). However, as a result of the standard pressure reduction method including the plateau correction, misleading sea level pressure values can be obtained for high-altitude stations (Mohr 2004). Evaluation of the existing pressure reduction method and correction of the pressure reduction problem are beyond the scope of this paper. We just need to be aware of this problem and to set a larger tolerance of R_z for highly elevated stations according to the mean and standard deviation of its R_z time series, so that other types of errors can still be identified.

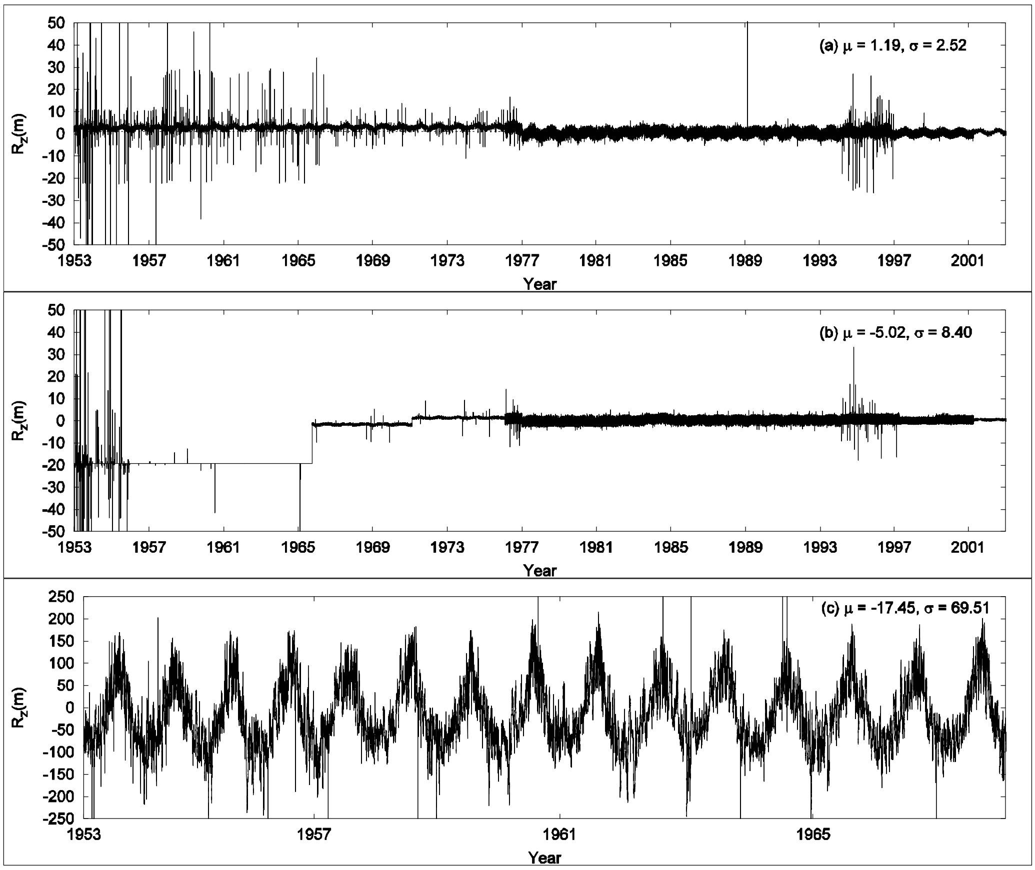


Figure 2. Time series of differences between the recorded and the estimated station elevation for hourly observations at (a) Abbotsford Airport, British Columbia (BC); (b) Victoria Int'l Airport, BC; and (c) Old Glory Mountain, BC.

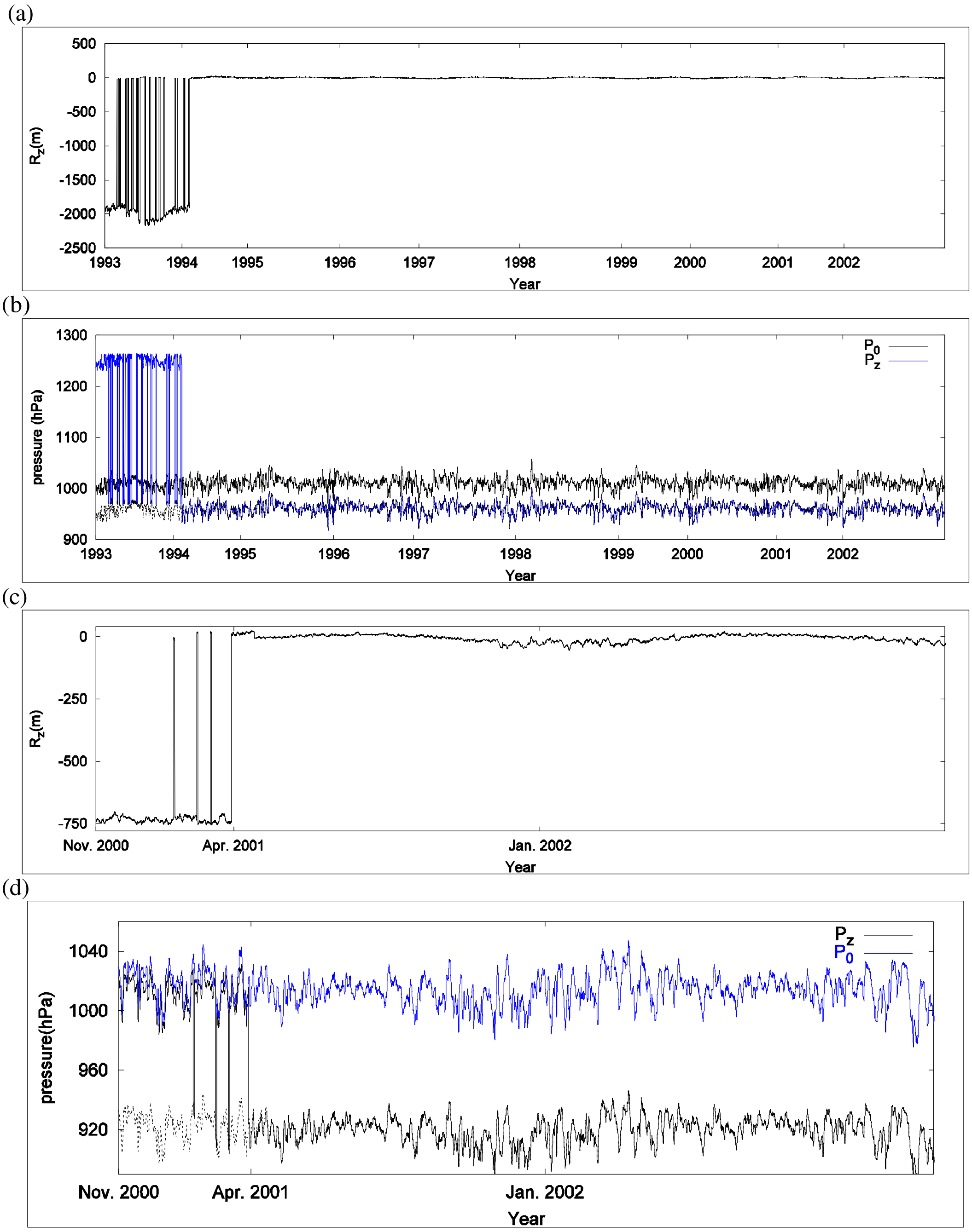
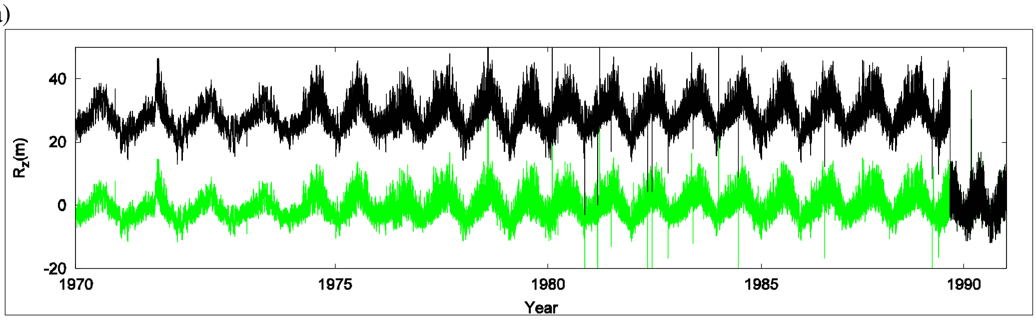


Figure 3. Time series of R_z, P_0, and P_z for the selected period of hourly observations at (a, b) Cape Hooper, NU and (c, d) Dease Lake LWOS, BC. The dash curve in b and d shows the corrected P_z values.

Figure 3 shows another type of error that was found for many stations, especially the Arctic stations. For some reason (maybe an error in the archive data ingestion), the station pressure values for the period from 1992-2002 were wrongly loaded for 40 stations, including 18 Arctic stations. Although it is hard to find the exact reason that caused this kind of error, the associated R_z values (Figs. 3a and 3c) are incredibly high, showing a clear step-change that would be easy to detect statistically, and most of the associated P_z values (see Figs. 3b and 3d) are unrealistically high and obviously wrong. This type of errors can be easily identified through visualization of the R_z and pressure time series together, as shown in Fig. 3.

Data inhomogeneities caused by station relocation (with big elevation change), observing instrument change (e.g., sensor used in automatic stations), and so on could also lead to step-changes (mean-shifts) in the R_z time series and hence large μ and σ values. As shown in Fig. 4a, the R_z time series for station Lytton (BC) shows a clear step-change on 1 July 1989, which was found to have arisen from a station relocation with 27.4m elevation decrease. This step-change can also be identified from the original time series of P_z (Fig. 4b).

(a)



(b)

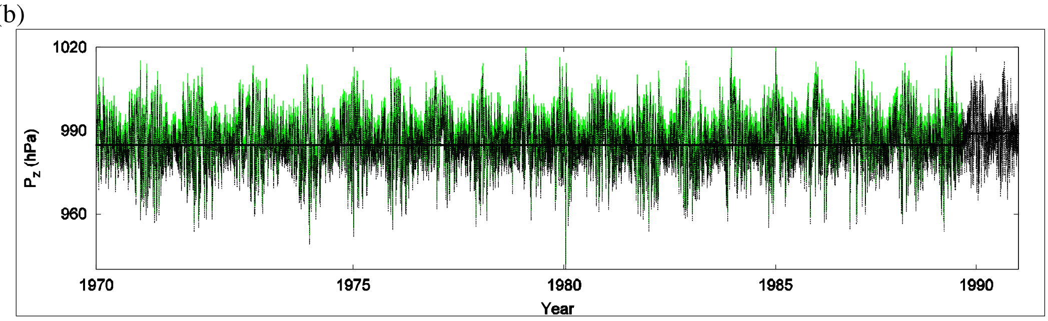


Figure 4. Time series of R_z (a) and P_z (b) for the selected period of hourly observations at Lytton (BC). The green curves in both panels indicate the adjusted values. The bold line in b shows the mean value of raw P_z before and after the change-point.

Since step-changes in the R_z time series could affect the estimates of μ and σ and hence the R_z tolerance used in hydrostatic check, we need to identify, and correct for, significant step-changes in R_z time series so that more realistic R_z tolerance can be determined. Considering the nature of the R_z time series, for the vast majority of stations it is reasonable to assume that R_z has an independent identical Gaussian distribution with mean μ and variance σ^2 under the null hypothesis of no step-changes. Thus, testing whether or not there is a step-change in the R_z time series for the period from N_1 to N_2 ($1 \le N_1 < N_2 \le N$; $n = N_2 - N_1 + 1$) is to test

$$H_0 : R_z(t) = \mu + \varepsilon_t \tag{5}$$

against

$$H_a : R_z(t) = \begin{cases} \mu_1 + \varepsilon_t, & N_1 \le t \le c \\ \mu_2 + \varepsilon_t, & c \le t \le N_2 \end{cases} \tag{6}$$

where step-size $\Delta = \mu_2 - \mu_1 \ne 0$ and ε_t denotes a zero-mean Gaussian variable with variance σ^2. In the first homogeneity test (i.e., at the beginning of the process), $N_1 = 1$ and $N_2 = N$. In the successive tests, N_1 or N_2 or both of them are set to the changepoints identified in the previous test(s), that is, the time series is segmented at the newly identified changepoint and a successive test is applied to each new segment of the time series (N_1 and N_2 are the first and last data point of the segment being tested; see Wang and Feng 2004 for the details).

Detection of an undocumented step-change can be done with the $T_{\max}$ statistic as in the Standard Normal Homogeneity test (Alexandersson 1986), or equivalently using the following $F_{\max}$ statistic:

$$F_{\max} = \max_{N_1 \le c \le N_2} F_c \tag{7}$$

where

$$F_c = \frac{(SSE_0 - SSE_a)/1}{SSE_a/(n-2)} \tag{8}$$

and

$$\begin{aligned} SSE_0 &= \sum_{t=N_1}^{N_2} \left[R_z(t) - \hat{\mu}\right]^2 \\ SSE_a &= \sum_{t=N_1}^{c} [R_z(t) - \hat{\mu}_1]^2 + \sum_{t=c+1}^{N_2} [R_z(t) - \hat{\mu}_2]^2 \end{aligned} \tag{9}$$

Similar to those in Wang (2003) and Lund and Reeves (2002), the critical values of the $F_{\max}$ statistic here are obtained from 10 million simulations under H_0 for each series length n. Also, the F_c statistic above, which has an F-distribution with (1, n-2) degrees of freedom under H_0, can be used to assess significance of a documented step-change (i.e., one that is supported by metadata) at time c. Both the $F_{\max}$ and F_c statistics are used in this study to identify R_z time series that have a significant step-change, along with metadata (if available). These R_z time series are further investigated, along with the related P_z and P_0 time series, to identify the cause and correct for the step-change (via correcting the erroneous pressure values).

(a)

(b)

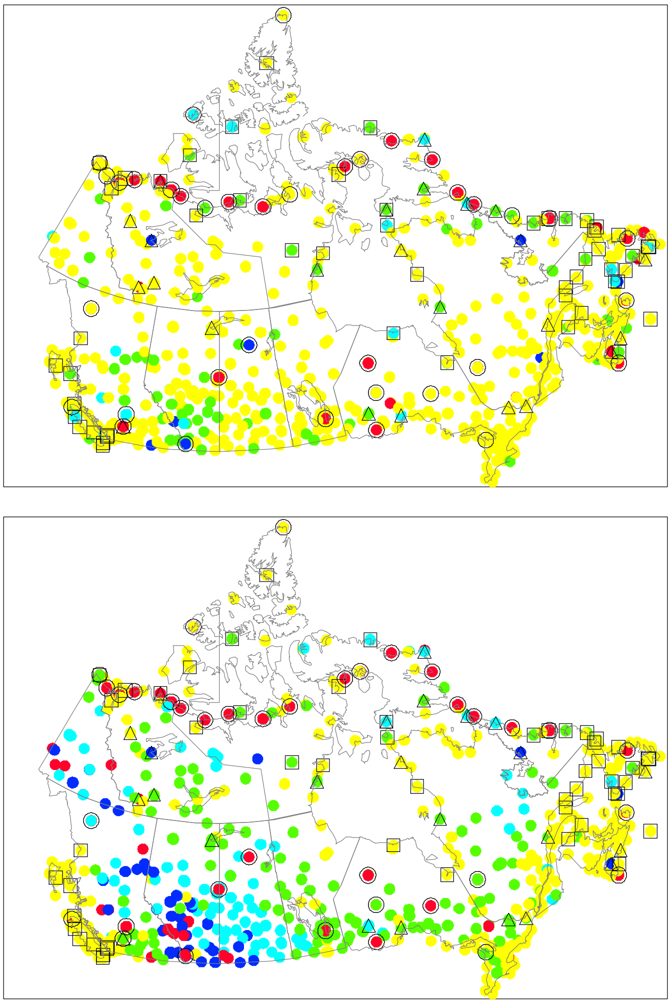


Figure 5. Absolute values of the mean and standard deviation of R_z (unit: m) time series calculated from raw pressure data. Stations of large step-changed(s) in the R_z time series are marked with a square to indicate the cause being the "50 feet rule" problem, a circle to indicate a long run of obviously wrong P_z values, and a triangle to indicate a station relocation without updating the station elevation.

As shown in Figure 5, significant step-change(s) in the R_z time series are found to have mainly arisen from either the "50 feet rule" problem (see those marked with a square), or a long run of obviously wrong P_z values (see those marked with a circle), or a station relocation without updating the changed elevation (see those marked with a triangle). The absolute values of the mean and standard deviation of R_z time series calculated from raw (uncorrected) pressure data, which are also shown in Figure 5, are large at the stations of obviously wrong P_z values (probably due to erroneous data ingestion), much larger than those at other problematic stations (because the error in elevation due to the "50 feet rule" problem or station relocation is relatively small).

Once all the R_z time series are corrected for all step-changes identified, their new mean and standard deviation (calculated from corrected pressure data) can be used in Eq. (4) to set the R_z tolerance, in which the γ value can now be selected by predetermining the rate of error using the methodology mentioned in Hubbard et al. (2005). In this study, we want to cap the random error rate uniformly across the country. We select the values of γ with the goal of keeping a 0.2‰ error rate for each station (thus, for a station with 50-year hourly observations, there will be 87 data flagged for further investigation).

e. Temporal consistency check (TC)

If a constant pressure value runs consecutively for 12 hours or longer in duration, all these hours are flagged as a result of the temporal consistency check.

4. DECISION MAKING METHOD (DMM)

We apply the afore-described five checks to hourly station and sea level pressure data (P_z and P_o) recorded at each station, subsequently. As a result, many values could be flagged in one or several or all of the five checks. However, not all flagged values are erroneous data. For example, a value can be flagged because of an error in the value recorded 1-3 hours earlier or later that cause the related pressure change to exceed its limit. One needs to analyze adjacent flagged values and the number of flags on each value, to find out the most suspicious one(s) for correction or exclusion. Such an analysis also leads to the removal of flags on values that are deemed correct. Thus, this decision making procedure is an important step in climate data quality assurance. Since the QA system is only applied to two elements, the decision making system is not very complicated. For example, a station pressure of 1006.4 hPa at 00:00 of 4 April 1954 was miscoded as 1016.4 hPa, which caused 11 flags as shown in Table 1. Usually a datum with the highest count of flags is most suspicious, and all flags on values adjacent to that datum can often be removed (e.g., the value 1016.4 is flagged in the final database and all other data in Table 1 are cleared of flags). This is the base of our automatic decision making method (DMM).

Table 1. Station pressure (P_z) and sea level pressure (P_o) recorded at Nanaimo, BC from 21:00Z of 3 April 1954 to 03:00Z of 4 April 1954, and the results of applying the five checks on these data. The LC, PC, HC, TC, and IC stand for the limit check, the pressure change check, the hydrostatic check, the temporal consistency check and the internal consistency check, respectively.

	p_z / p_0 (hPa)	LC flag (p_z / p_0)	PC flag (p_z / p_0)	HC flag (p_z / p_0)	TC flag (p_z / p_0)	IC flag (p_z / p_0)	*Total flags* (p_z / p_0)
21:00	1006.8/1010.6	0/0	1/0	0/0	0/0	0/0	1/0
22:00	1006.8/1010.6	0/0	1/0	0/0	0/0	0/0	1/0
23:00	1006.7/1010.5	0/0	1/0	0/0	0/0	0/0	1/0
0:00	**1016.4/**1010.2	0/0	1/0	1/1	0/0	1/1	3/2
1:00	1005.7/1009.5	0/0	1/0	0/0	0/0	0/0	1/0
2:00	1004.8/1008.6	0/0	1/0	0/0	0/0	0/0	1/0
3:00	1003.9/1007.7	0/0	1/0	0/0	0/0	0/0	1/0

Occasionally, the total counts of flags for the two elements (P_z and P_O) are the same and we do not have enough information to judge which element is more suspicious. For example, a valid P_O of 1021.5 hPa is miscoded as 1025.1 hPa, which is a mild error, not severe enough to raise the LC/PC/TC/IC flags, only enough to raise the HC flags. In this case, we can not determine which element (P_z or P_O) is erroneous; thus, both the P_z and P_O values are flagged and further inspected manually.

5. CORRECTION OF ERRORS

Errors in meteorological data are very complicated and not easy to correct. Nevertheless, we should try our best not to reject, but to be able to correct erroneous data, especially for data-sparse regions. An automatic error-correction system is designed in the study.

It is highly desirable to know what caused the errors before we start to correct them. Table 2 lists the four types of errors that are most often found in our digital hourly pressure database, in addition to those that lead to a significant step-change in the R_z time series. The vast majority of errors are of Types 1 and 2 (see Table 2). The Type 3 error is a profound problem in the Canadian hourly pressure data that were digitized from paper archives. In Canada, hourly pressure values used to be recorded

Table 2. Errors most often found in the Canadian digital pressure database.

Type	Description
1	One digit is miscoded (e.g., 1 is mistaken for 0, 2 for 3, 1 for 7, etc.)
2	Digits are transposed (e.g., 1032.5 entered as 1035.2)
3	Wrong base number added (e.g., “73” is taken as 907.3 hPa when it should be 1007.3 hPa)
4	Station pressure and sea level pressure are transposed or have the same value when the station elevation $Z \neq 0$.

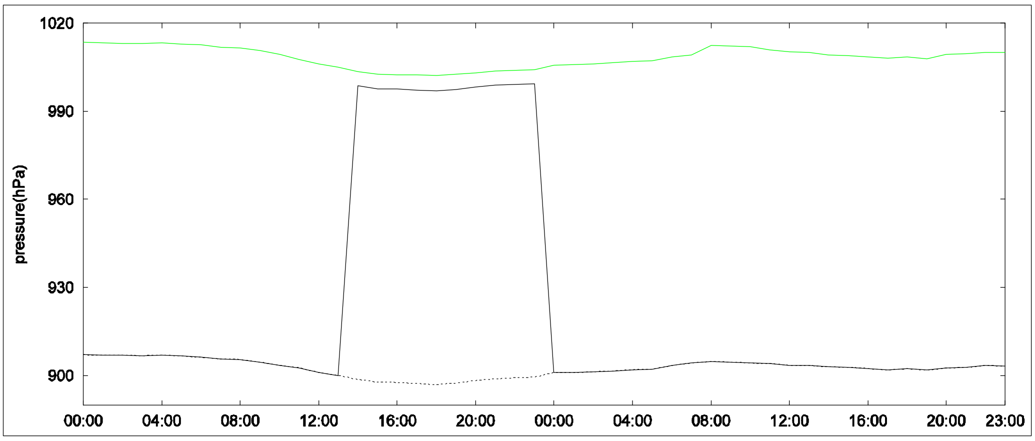


Figure 6. An example of using a wrong base number when digitizing station pressure data recorded at Red Deer Airport (Alberta) from 0:00 on April 20 to 23:00 on April 21 1953. The dashed line shows the correct values, and green line shows the corresponding sea level pressure values.

(manually on paper) in tenths of hPa and only the last three digits were recorded (e.g., “132” for a pressure of 10132, or “587” for 9587; unit: 0.1 hPa). The omitted base number (10,000 or 9000, or even 8000) needs to be added back during the digitization of our paper archives. Unfortunately, it is not always easy to determine which base number should be added, and the algorithm used to do so makes mistakes. This is why this type of errors occur and can be very hard (even impossible) to correct. This type of errors sometimes persists for several hours or days, or even months (cf. Figure 6) and can be mistaken as systematic biases caused by station relocation or instrument change, etc. Unfortunately, the same base number problem affects the NCEP/NCAR reanalysis dataset for the period from 1948-67 (NCEP/NCAR, 2006). Such errors will only be detected by the hydrostatic or internal consistency check (usually they will not exceed the climatological limits); thus, it is sometimes impossible for us to determine whether P_z or P_o is in error. A visual inspection of the time series segment often helps identify this type of errors, which we do in this study.

a. Correction of systematic errors

As mentioned in section 3d, the hydrostatic check is useful in identifying and correcting systematic errors that lead to a significant step-change in R_z time series, such as those caused by the "50 feet rule" problem, by a long run of obviously wrong P_z values (e.g., those shown in Fig. 3), and by station relocation without updating the changed elevation. We found that all the systematic step-changes in R_z time series are associated with erroneous P_z (but correct P_o) values. Correction of this kind of systematic errors is relatively straightforward. These systematic errors have one common feature, that is, they are due to a change/error in elevation Z. Theoretically, we can simply use the correct station elevation and Eq. (10b) to calculate the correct P_z values and use them to replace the corresponding erroneous P_z values. However, stations of these systematic errors could be in the elevated areas (except for those of the "50 feet rule problems) and hence their R_z time series could have large periodic variations such as those shown in Figure 4a (which are due to the elevated area pressure reduction problem; see discussions in section 3d). Replacement of erroneous P_z values with the corresponding P_z values calculated using the correct elevation would completely remove the periodic feature of the R_z time series (forcing zero R_z values throughout the period of correction), which is not desired here. In this case, the desirable correction is the difference between the mean (over the period of wrong elevation) of the calculated P_z values (say $\overline{P}_z^c$) and that of the erroneous P_z values (say $\overline{P}_z^e$), that is, we just need to add $\Delta = \overline{P}_z^c - \overline{P}_z^e$ on the erroneous P_z values to obtain the corrected P_z values. For example, for the Lytton case shown in Fig. 4, we add $\Delta = 3.4\,\text{hPa}$ to all the P_z values before 1 July 1989. Such a correction only corrects for the systematic error, without changing the peculiar feature of R_z time series (see Fig. 4a). Of course, random errors are still to be identified and corrected for. Corrections of random errors are described below.

b. Correction of isolated simple errors

Errors of the Type 1 or 2 (see Table 2) are usually isolated cases (i.e., the values before and after it are correct for both elements) that are easy to correct and hence called simple errors. The algorithm we use to correct an isolated case of simple

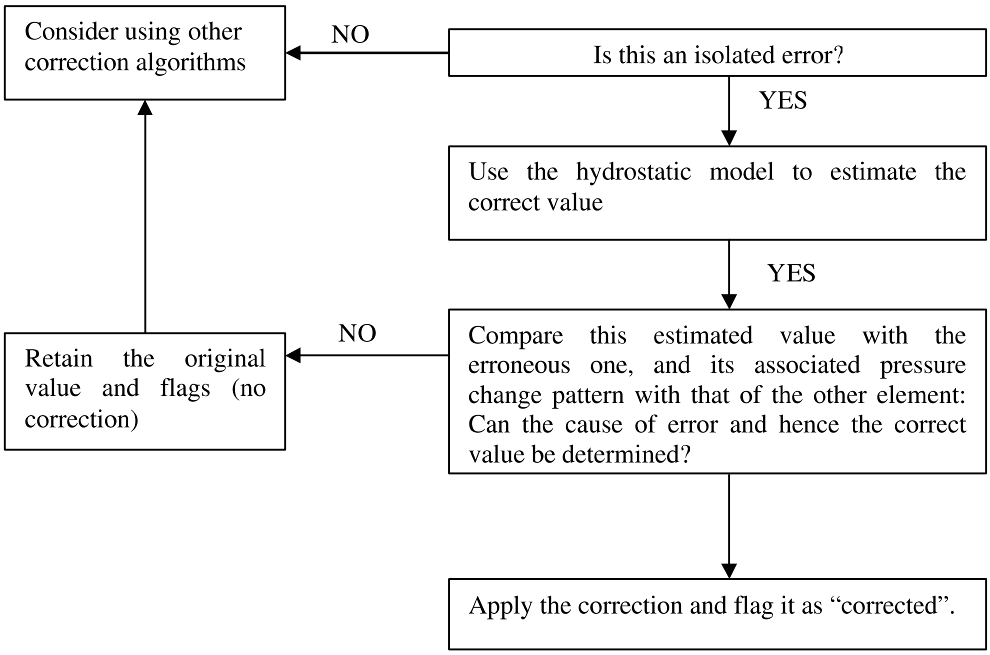


Figure 7. An algorithm for correcting isolated simple errors.

error is outlined in Figure 7. First, we determine if the erroneous datum is an isolated error. If yes, we use the hydrostatic model to estimate the correct value, that is,

$$P_o = \exp\left(\frac{(g/R)\times Z}{T_{mv}}\right)\times P_z \tag{10a}$$

or

$$P_z = \frac{P_o}{\exp\left(\frac{(g/R)\times Z}{T_{mv}}\right)} \tag{10b}$$

depending on which element is in error. We use the recorded station elevation here, and a plateau correction was added in T_{mv} if the error occurs after November 1976 (the time the plateau correction was introduced in Canada), using the plateau correction parameters taken from the EC archive. Then, we compare this estimated pressure value with the original (erroneous) one, and compare its associated pressure change pattern with the corresponding pattern of the other element (the two elements should have the same pattern of pressure change), to see if we can determine the cause of error and hence the correct value. As shown in Tables 3 and 4, if replacement of a digit or a transposition of two digits in the original data would make it approximately equal to the estimated value and ensure a consistency of pressure change between the two elements, this is a simple error; we apply the correction and flag it as "corrected". If this is not a simple error and we are not able to determine the cause or the correct value, or if this is not an isolated case of error, we consider using other error-correction algorithms (see the next subsections).

Table 3. An example of the Type 1 error: "1029.1" was mis-keyed in as "1024.1" (1953/01/16 at station 7016294).

	20:00	21:00	22:00	model value	correct value
P_o	1027.3	**1024.1**	1029.9	1029.1	**1029.1**
P_z	1017.6	1019.4	1020.2		

Table 4. An example of the Type 2 error: "59.2" was mis-keyed in as "52.9" (1965/11/25 at station 4019080).

	06:00	07:00	08:00	model value	correct value
P_z	959.5	**952.9**	958.6	957.4	**959.2**
P_o	1024.7	1024.1	1023.5		

c. Correction of isolated but non-simple errors

Sometimes, an isolated error is not a simple error (of the Type 1 or 2). For example, the value 846.6 in Table 5 is completely wrong, inconsistent with either the corresponding or neighboring hourly P_z or P_o values. The hydrostatic model estimate of the correct value is 1023.8, which would ensure a consistent pressure change pattern for both elements here and would pass the pressure limit check if it were used to replace the erroneous value 846.6. In other words, it is reasonable to replace 846.6 with 1023.8 in this case. Thus, we apply the correction and flag it as "corrected".

Table 5: An example of more than two digits in error (1976/02/04 at station 1018642).

	08:00	09:00	10:00	model value	correction
P_z	1023.5	**846.6**	1024.0	1023.8	**1023.8**
P_o	1027.4	1027.8	1027.9		

d. Human-machine interactive corrections

The existing QA methods are often not able to correct erroneous data completely automatically. Human-machine interactive correction is usually applied when the automatic decision making method can not determine which element is in error. In this case, one needs to analyze manually the flag types and the original data for both elements, to determine which element is in error, and to estimate the correct value (s). In most cases, the correction is set to the value estimated using the hydrostatic model. For example, our analysis of the data shown in Table 6 reveals that the P_z value of 1001.6 was mistaken as the P_0 value, whose reasonable estimate is 996.7.

Table 6. An example of mistakenly reporting the same value for both sea level pressure P_o and station pressure P_z (1954/10/28 at station 7113534).

	19:00	22:00	1:00	model value	correction
P_z	996.3	**1001.6**	998.2	996.7	**996.7**
P_0	1000.9	1001.6	1002.8		

Actually, human-machine interaction was involved in the correction of those systematic errors described in section 5a. The corresponding R_z time series is visualized, along with both P_o and P_z time series to determine the error and its cause, as shown earlier in Figure 3, because the automatic decision making system is not able to determine which element (P_o or P_z) is in error in this case, although the hydrostatic check is powerful in identifying and correcting this type of errors.

Finally, there exist a very small number of suspicious reports that even a specialist was not be able to correct. This situation usually occurs when the hydrostatic check can not be performed because of a missing element (e.g., dry bulb temperature) that is needed as input to the hydrostatic model. In this case, we set the data as missing if they do not pass the climatological limits check. Otherwise, we accept them without any correction.

6. ANALYSIS OF THE CORRECTED DATA SERIES

The QA approach described above is applied to each station for both pressure levels. Corrected data are stored with their corresponding flags. However, a second iteration of QA was run with corrected data in order to detect any wrong correction or erroneous data that went undetected at the first run.

Figure 8 shows the rate of random errors identified for each station (systematic errors that were corrected as described in section 5a were not counted here). However, the error rates for most stations are under 1‰. Of more than 1.8×10^8 hourly pressure data (both level) processed, approximately 4.1×10^6 (or 2.3%) data (including systematic errors) have been corrected. About 30% of those detected errors can be automatically corrected, while human-machine interactive correction is needed to correct the other 70%.

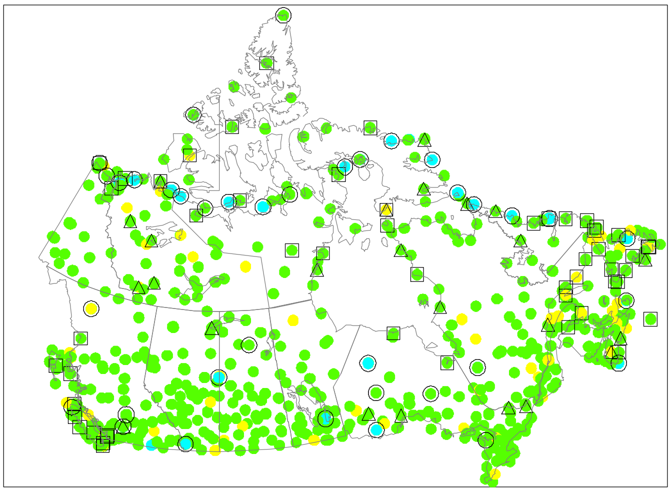


Figure 8. The rate of random errors identified/corrected for each station (unit: ‰). The symbols on stations are the same as in Figure 5.

As shown in Fig. 9, the standard deviation of the R_z times series calculated using corrected station and mean sea level pressure data are much smaller, showing a better organized pattern. Large values are now seen only at the elevated stations.

The hydrostatic check plays an important role in the whole QA system. About 50% of errors were detected and corrected through this check. Also, our results show that it is reasonable to assume the correctness of the hourly dry bulb temperature data in the hydrostatic check. Actually, the hydrostatic method can also be helpful in detecting inhomogeneities in atmospheric pressure data caused by station relocation, observer change etc. as shown in Fig. 4.

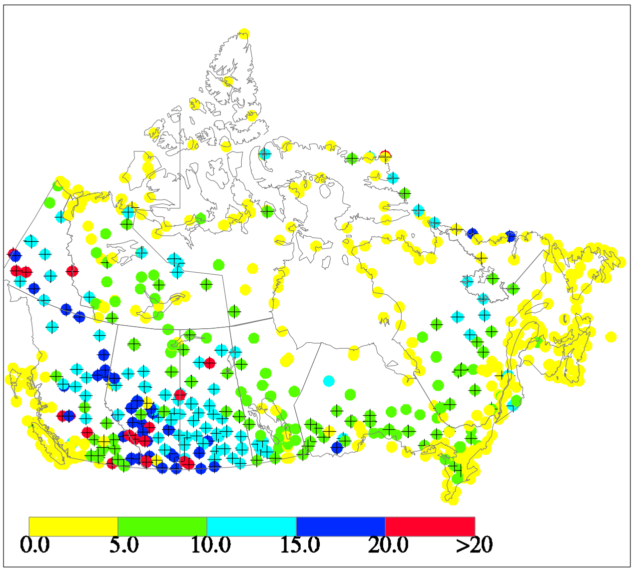


Figure 9. The same as in Fig. 5b but for those calculated from corrected pressure data. The plus sign indicates that the station elevation is greater than 305m.

7. CONCLUDING REMARKS

Aiming to build a high quality dataset for both station and sea level pressure in Canada, we have developed a comprehensive QA system for surface atmospheric pressure at two levels (station- and sea-level), which was applied to pressure data recorded in the last 50 years at 761 Canadian stations.

The results show that there exist serious systematic errors in the Canadian historical atmospheric pressure data and that random error(s) can be found for almost every station. The systematic errors are found to be caused either by the use of wrong station elevation values in the reduction of barometer readings to station or sea level pressure values (e.g., the "50 feet rule" or station relocation without updating the station elevation), or by transposition/swapping of station and sea level pressure values, or by mistakes made in the archive data ingestion or data recording/digitization processes (e.g., use of a wrong base number). Fortunately, a vast majority of these errors can be detected and corrected by the QA system with either automatic or interactive correcting method. The corrected P_o and P_z data should be much more reliable and better suited for various climate studies, including its use in producing a 100-yr reanalysis (Compo et al., 2006).

Acknowledgements

The authors wish to thank Dr. Gilbert Compo and Mr. Amir Shabbar for their helpful comments/suggestions on an earlier version of this manuscript.

References

Alexander, L., S. Tett, and T. Jonsson, 2005: Recent observe changes in severe storms over the United Kingdom and Iceland. Geophys. Res. Lett., 32, L13704, doi: 10.1029/2005GL022371.

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. J. of Climatology, 6: 661-675

Allan R. J., J. A. Lindesay, and D. E. Parker, 1996: El Nino southern oscillation and climatic variability. CSIRO Publications, 405pp.

Christensen, R. A. and R. F. Eilbert, 1985: Seasonal precipitation forecasting with a 6-7 month lead time in the Pacific Northwest using an information theoretic model. Mon. Wea. Rev., 113: 502-518.

Collins, W. G. and L. S. Gandin, 1990: Comprehensive hydrostatic quality control at the National Meteorological Center. Mon. Wea. Rev., 118: 2752 – 2767.

Compo, G. P., J. S. Whitaker, and P. D. Sardeshmukh, 2006: Feasibility of a 100 year reanalysis using only surface pressure data. Bull. Amer. Met. Soc., 87, 175-190.

Environment Canada, 2004: The National Archives System QC Program Checks for Hourly SA's. Available online at http://nadm.ontario.int.ec.gc.ca/Intranet/nadm/nas/quality_control/hourlies_e.html (accessible within Environment Canada only)

Environment Canada, 1977: Manual of Surface Weather Observations, Seventh Edition.

Gandin, L. S., 1988: Complex quality control of meteorological observations. Mon. Wea. Rev., 116: 1137 – 1156.

Graybeal, D. Y., A. T. DeGaetano, and K. L. Eggleston, 2004: Complex quality assurance of historical hourly surface airways meteorological data. J. of Atmos. and Ocean. Tech., 21: 1156 – 1169.

Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.), 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 881pp

Hubbard, K. G., S. Goddard, W. D. Sorensen, N. Wells and T. T. Osugi, 2005: Performance of quality assurance procedures for an applied climate information system. J. of Atmos. and Ocean. Tech., 22: 105 – 112.

Jones, P. D., M. J. Salinger, and A. B. Mullan, 1999: Extratropical circulation indices in the southern Hemisphere based on station data. Int. J. of Climatol., 19: 1301-1317.

Kaplan, A, Y. Kushnir, M. A. Cane, 2000: Reduced space optimal interpolation of historical marine sea level pressure: 1854 – 1992. J. of Climate, 13: 2987 – 3002.

Kunkel, K. E., and Coauthors, 1998: An expanded digital daily database for climatic resources applications in the Midwestern United States. Bull. Amer. Meteor. Soc., 79, 1357-1366.

Le Blancq, F. W., 2003: Newfoundland low and pressure tendencies, 11 February 2003. Weather, 58: 239 – 242.

Lund, R. and J. Reeves, 2002: Detection of undocumented changepoints: A revision of the two-phase regression model. J. of Climate, 15: 2547-2554.

Mohr, M., 2004: Problems with the mean sea level pressure field over the western United States. Mon. Wea. Rev., 132: 1952 – 1965.

NCEP/NCAR, cited 2006: NCEP/NCAR Reanalysis PSFC problem 1948-1967. Available online at http://wwwt.emc.ncep.noaa.gov/gmb/bkistler/psfc/psfc.html.

Nkemdirim, L. C., and D. Budikova, 2001: Trends in sea level pressure across western Canada. J. of Geophys. Res., D 1061: 11801 – 11812.

Pauley, P. M., 1998: An example of uncertainty in sea level pressure reduction. Weather and Forecasting, 13: 833 – 850.

Savdie, I. 1982: AES barometry program, Technical Record No. 9. Network Planning and Standards Division. Data Acquisition Services Branch, Environment Canada, Downsview, Ontario, Canada.

Shafer, M. A., C. A. Fiebrich, D. S. Arndt, S. E. Frederickson, and T. W. Hughes, 2000: Quality assurance procedures in the Oklahoma Mesonetwork. J. of Atmos. and Ocean. Tech., 17: 474 – 494.

Slonosky, V. C., and E. Graham, 2005: Canadian Pressure Observations and Circulation variability: Links to Air Temperature. Int. J. of climatol., 25: 1473 – 1492.

Smith, T. M., and R. W. Reynolds, 2003: Reconstruction of Monthly Mean Oceanic Sea Level Pressure Based on COADS and Station Data (1854 – 1997). J. of Atmos. and Ocean. Tech., 21: 1272 – 1282.

Trenberth, K. E., and D. A. Paolino, 1980: The Northern Hemisphere sea-level pressure data set: Trends, errors and discontinuities. Mon. Wea. Rev., 108: 855 – 872.

Wang, X. L., H. Wan, and V. R. Swail, 2006: Observed changes in cyclone activity in Canada and their relationships to major circulation regimes. J. Climate, **19** (No. 6), 896–915.

Wang, X. L. and Y. Feng, cited 2004: RHTest User Manual. Available online at http://cccma.seos.uvic.ca/ETCCDMI/RHTest/RHTestUserManual.doc.

Wang, X. L., 2003: Comments on "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model". J. of Climate, 16, 3383-3385.

WMO, 1954: Reduction of Atmospheric Pressure. WMO Technical Note No. 7.

WMO, 1964: Note on the standardization of Pressure Reduction Methods in the International Network of Synoptic Stations. WMO Technical Note No. 61.

Wright, P. B., 1984: Relationships between indices of the Southern Oscillation. Mon. Wea. Rev., 112: 1913 – 1919.

A recursive testing algorithm for detecting and adjusting for multiple artificial changepoints in a time series

Xiaolan L. Wang
Climate Research Division, ASTD, STB, Environment Canada, Toronto, ON, Canada
Phone: +1-416-739-4115 Email:Xiaolan.Wang@ec.gc.ca

1. INTRODUCTION

This algorithm is based on one of the following "two-phase regression" (TPR) approaches for detecting a changepoint at time *c* in time series X_t $(t = 1,2,...,N)$:

- TPR2 is to test

$$H_0 : X_t = \mu + \varepsilon_t \quad \text{(null model)} \quad \text{against}$$
$$H_a : X_t = \mu + \Delta I_{[t>c]} + \varepsilon_t \quad \text{(full model)}$$

  [note that TPR2 with an independent, identically distributed (IID) Gaussian noise process is equivalent to the SNHT in Alexandersson (1986)]

- TPR3 is to test

$$H_0 : X_t = \mu + \beta t + \varepsilon_t \quad \text{(null model)} \quad \text{against}$$
$$H_a : X_t = \mu + \Delta I_{[t>c]} + \beta t + \varepsilon_t \quad \text{(full model)}$$

- TPR4 is to test

$$H_0 : X_t = \mu + \beta t + \varepsilon_t \quad \text{(null model)} \quad \text{against}$$
$$H_a : X_t = \mu + \Delta I_{[t>c]} + (\beta + \delta I_{[t>c]})t + \varepsilon_t \quad \text{(full model)}$$

- TPR3[a] is to test

$$H_0 : X_t = \mu + \beta t + \varepsilon_t \quad \text{(null model)} \quad \text{against}$$
$$H_a : X_t = \mu - \delta c I_{[t>c]} + (\beta + \delta I_{[t>c]})t + \varepsilon_t \quad \text{(full model)}$$

  [this is a special case of TPR4, in which $\Delta = -\delta c$ (i.e., no mean-shift, only the trend changes at time c); so the full model has only 3 free parameters]

where

$$I_{[t>c]} = \begin{cases} 0 & \text{for } t \le c \\ 1 & \text{for } t > c \end{cases}$$

and ε_t can be *either* an IID Gaussian noise process (Reeves *et al*. 2006, Wang 2003, Lund and Reeves 2002) *or* a periodic Gaussian noise process (with periodic lag-1 autocorrelation and perhaps also periodic variance; see Lund *et al*. 2006 for details). For an <u>undocumented changepoint</u>, the test statistic is

$$F_{\max} = \max_{1 \le c \le N-1} F_c$$

where

$$F_c = \frac{(SSE_0 - SSE_a)/(m_a - m_0)}{SSEa/(N - m_a)},$$

m_0 and m_a are the number of free parameters involved in the null model and the full model, respectively, and SSE_0 and SSE_a are the sum of squared errors (SSE) of the null and full model fit (i.e., without and with a changepoint at time *c*), respectively. In the case of autocorrelated and/or periodic Gaussian noise, both the null and full model's residuals are scaled by the variance of the full model's residuals before being used to calculate the SSE's (see Lund *et al.* 2006 for details). The significance of an undocumented changepoint is determined by comparing its $F_{\max}$ value with the corresponding 95th percentiles F_{m05} (Reeves *et al.* 2006, Lund *et al.* 2006, Wang 2003, Lund and Reeves 2002). **However**, the significance of a documented changepoint (say at a known time c_0) is determined by comparing its F_{c_0} value with the F-distribution with $(m_a - m_0)$ numerator degrees of freedom and $(N - m_a)$ denominator degrees of freedom (its significance level α is estimated).

2. THE RECURSIVE TESTING ALGORITHM

A recursive testing procedure is necessary, because of the possible existence of multiple change-points in one single time series, and because most of the current change-point detection methods are developed assuming a single change-point in the time series. As a result, the first change-point identified might be false or inaccurate, due to "contamination" by other change-points in the time series.

This recursive testing algorithm (see also Figure 1) goes like this:

[1] Set $I_x = 0$ (indicating more changepoints can be added to the list), $I_f = 0$ (indicating this is not final assessment), and $I_d(k) = 1$ for $k = 1,2,..., N-1$ (i.e., assume that all possible changepoints are documented in the screening process prior to the final assessment; this will be verified later in [4]). As in the conventional hierarchical (or step-wise) splitting algorithm, the first check is done on the whole series, to find out the time c_x that is associated with $F_{c_x} = F_{\max} = \max_{1 \le c \le N-1} F_c$. Then, the time series being tested is split into two segments, $S_1 \in [1,2,..., c_x]$ and $S_2 \in [c_x + 1, c_x + 2, ... N]$, regardless of the significance of the possible changepoint at time c_x (the significance is to be re-assessed later in [3]). Set $N_c = 1$, $c_1 = c_x$, the list of

changepoints $\vec{C}=\{c_1\}$, and $I_{all}=0$ (indicating **not all** the listed changepoints are significant). Save $SSE_a(N_c)=SSE$ of the full model with a changepoint at time $c_1=c_x$.

[2] For each possible value of $c\in\Re$ ($\Re$ includes all integers in interval $[1, N-1]$ except the N_c points already listed as changepoints), fit a (N_c+2)-phase regression with the N_c changepoints **plus** a new candidate at time *c*, and save the resulting *SSE* as $SSE(c)$ (note that these are the full model's *SSE*). Then, find the new changepoint c_x by searching for $SSE(c_x)=\min_{c\in\Re} SSE(c)$; $F_{\max}=F_{c_x}=\dfrac{SSE_a(N_c)-SSE(c_x)}{SSE(c_x)/(N-m)}$, where *m* is the number of free parameters involved in the full model with the $M_c=(N_c+1)$ changepoints (e.g., $m=(2+M_c)$ for the IID TPR3 model, and $m=(1+M_c)$ for the IID SNHT model). **If** the changepoint c_x **is significant** in the presence of the other N_c changepoints [i.e., it has $\alpha\le 0.05$ if $I_d(c_x)=1$ (documented) or $F_{\max}\ge F_{m05}$ if $I_d(c_x)=0$ (undocumented)], add it to the list of changepoints to form a new list $\vec{C}=\{c_1<c_2<\cdots<c_{N_c+1}\}$, set $M_c=(N_c+1)$, $SSEa(M_c)=SSE(c_x)$, and $N_c=M_c$; if $I_{all}=1$ (i.e., all the listed changepoints are significant) go to repeat [2], otherwise go to [3]. **If** the changepoint c_x **is not significant**, set $I_x=1$ (no more changepoint can be added to the list) and $M_c=N_c$, then go to [3].

[3] For each *k* ($k=1,2,\ldots,M_c$), fit a M_c-phase regression to the data, omitting changepoint c_k while keeping all the other (M_c-1) changepoints in the list of $\vec{C}=\{c_1<c_2<\cdots<c_{M_c}\}$, and save the resulting *SSE* as $SSE_0(k)$ and $F_{\max}(k)=\dfrac{SSE_0(k)-SSEa(M_c)}{SSEa(M_c)/(N-m)}$. Then, find the **least significant** changepoint c_{k*} by searching for $F_{\max}(k*)=\min_{1\le k\le M_c} F_{\max}(k)$.

If c_{k*} is **significant** in the presence of the other (M_c-1) changepoints (it has $\alpha\le 0.05$ if $I_d(c_{k*})=1$ or $F_{\max}\ge F_{m05}$ if $I_d(c_{k*})=0$), all the M_c changepoints are significant, set $I_{all}=1$, $N_c=M_c$; if $I_x=0$ go to repeat [2], otherwise go to [4].

If c_{k*} is **not significant** in the presence of the other (M_c-1)

changepoints (it has $\alpha \leq 0.05$ if $I_d(c_{k*}) = 1$ or $F_{max} \geq F_{m05}$ if $I_d(c_{k*}) = 0$), delete it from the list of changepoints and set $M_c = M_c - 1$, **then** set $SSEa(M_c) = SSE_0(k*)$; if $M_c > 0$ go to repeat [3], otherwise go to [4].

[4] Output the estimated parameters and statistics of the ($M_c + 1$)-phase regression model fit. If $M_c > 0$ and $I_f = 0$, analyze these results along with metadata (if available): For $k = 1,2,...M_c$, set $I_d(c_k) = 0$ **if** c_k is found to have no metadata support (undocumented). Delete from the list the changepoint of the **smallest** shift if it turns out to be an insignificant undocumented changepoint, set $M_c = M_c - 1$ and $I_f = 1$, and go to repeat [3]; **otherwise**, all changepoints in the list are significant; these results are the final estimates, which include the final estimates of the position, significance, and magnitude of all shifts identified [along with all other non-changepoint parameters (e.g., intercepts, slope(s)…]. Most importantly, the position, significance, and magnitude of each artificial shift are assessed in the presence of all other changepoints in the time series.

Note that during the recursive testing process **before reaching step [4] for the first time** (when $I_f = 0$), the significance level α is estimated for all changepoints as if they were documented and all changepoints with $\alpha \leq 0.05$ are kept in the list of changepoints for further analysis along with metadata (in step [4]). All changepoints with $\alpha \leq 0.05$ but $F_{max} < F_{m05}$ that are found later to have no metadata support (undocumented changepoints) are deleted from the list (one at a time) when we go to repeat [3] from [4].

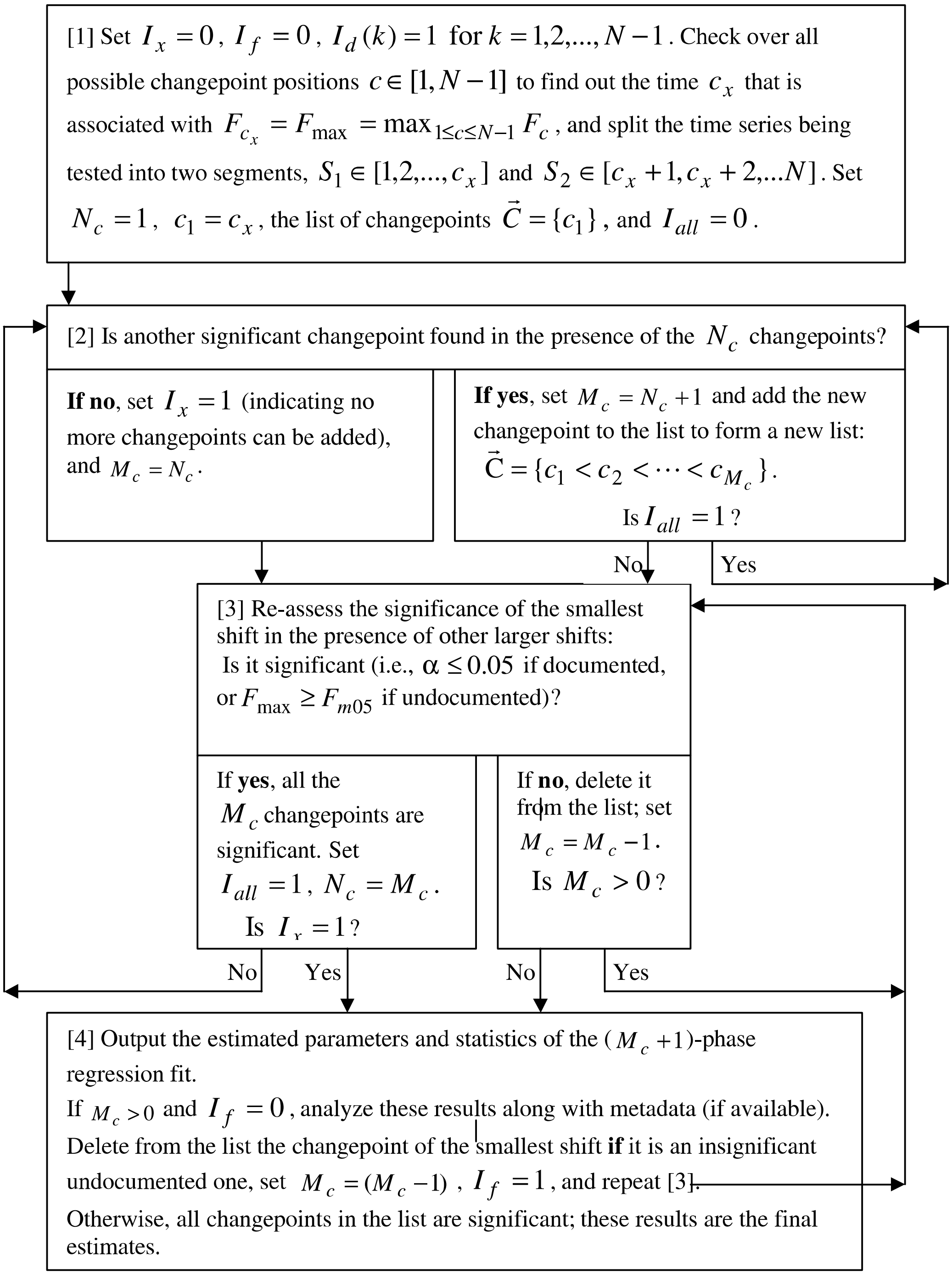


Figure 1. A recursive testing algorithm for detecting multiple artificial changepoints in a time series.

3. REMARKS ON MODEL SELECTION

The open source software package RHTest (Wang and Feng 2004), which is available free of charge at http://cccma.seos.uvic.ca/ETCCDMI/software.html (see also Wang 2006), has been modified to implement this recursive testing algorithm for detecting and adjusting for multiple artificial changepoints in a time series. Users of this software can choose an appropriate model (e.g., TPR2 or TPR3 or TPR4, with IID or periodic Gaussian noise) according to the characteristics of the time series to be tested.

Generally, TPR2 should be the best in cases of an available reference series that is good enough to completely remove the climate signal (trends and periodic fluctuations) from the target series (i.e., the target-minus-reference series has zero trend and no periodic fluctuations). TPR3 should be the best in cases of a systematic difference in trend between the reference and the target series. $TPR3^a$ should be used when only a gradual shift (no sudden change) is suspected in the target series (but not in the reference series if used). TPR4 is not recommended in most cases **when a reference series is available and used**, because TPR4 tends to overfit the series (Reeves *et al*. 2006) and, in particular, it is usually not realistic for the difference in trend between the reference and the target series to **change often over time** (in that case, the reference series is of little sense).

It should be stressed that the use of good reference series can not diminish the autocorrelation in the target series if exists. Thus, the TPR approaches with an autocorrelated noise process should be more appropriate than their IID counterparts in most climate applications. In particular, when a reasonably good reference series is not available, it is necessary to use a TPR approach that takes into account changepoints, autocorrelation, and periodicities (including seasonality) in tandem (see Lund et al. 2006). In this case, TPR2, with white or periodic noise, is often not suitable in climate applications because of the presence of climate signal in the target series; TPR3 with periodic mean response and periodic noise should be suitable for most applications, especially those with focus on identification of mean-shifts.

Also, it should be pointed out that homogeneity of reference series can not be assumed without some sort of investigation, and that relative homogeneity tests may let network-wide artificial shifts go undetected. Although large shifts can be seen through visual inspection of time series, it should be more objective to apply to the reference series a TPR approach that takes into account changepoints, autocorrelation, and periodicities in tandem, to obtain preliminary information about its homogeneity, to help determine if a changepoint in the target-minus-reference series comes from the target or from the reference series and if there is any network-wide artificial shift present in the observing network being analyzed.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. J. Climatol., 6, 661-675.

Lund, R., X. L. Wang, J. Reeves, Y. Feng, Q. Lu, C. Gallagher, 2006: Changepoint Detection in Periodic and Autocorrelated Time Series. *J. Climate* (submitted).

Reeves, J., J. Chen, X. L. Wang, R. Lund, Q. Lu, 2006: A Review and Comparison of Changepoint Detection Techniques for Climate Data. *J. App. Meteor. Climatol.*, accepted subject to minor revisions.

Wang, X. L., 2003: Comments on "Detection of Undocumented Changepoints: A Revision of the Two-Phase Regression Model". *J. Climate*, 16, 3383-3385. doi: 10.1175/1520-0442(2003)016<3383:CODOUC>2.0.CO;2.

Wang, X. L. and Y. Feng, 2004: *RHTest User Manual*. Available online at http://cccma.seos.uvic.ca/ETCCDMI/RHTest/RHTestUserManual.doc.

Wang, X. L., 2006: The RHTest package and future development. Included in this report.

PROGRAMME

Budapest, Hungary
29 May – 2 June 2006

<u>Venue:</u>

The Headquarters of the Hungarian Meteorological Service (1 Kitaibel Pál street, Budapest)

Monday, 29 May

8:00-10:00
Registration

10:00-12:00
Opening Addresses by

the President of HMS
the Organizers
Szentimrey, T. (HU): An overview on the main methodological questions of homogenization
Mestre, O. (FR): A review of homogenisation procedures

Lunch break

14:00-17:00
Mekis, É. and Vincent, L. (CA): Homogenization of precipitation, temperature and humidity in Canada: issues and recommendations
Lund, R., Wang, X.L., Reeves, J., Feng, Y., Lu, Q., Gallagher, C. (CA, US): Changepoint Detection in Periodic and Autocorrelated Time Series
Wang, X.L. (CA): The RHTest package and future development
Jourdain, S.J. (FR): Homogenisation of French long term data series
Marinova, T. and Alexandrov, V. (BG): On homogenization of Climate Long-term Series in Bulgaria
Stepanek, P. (CZ): Experiences with homogenization of monthly air temperature and precipitation series in the Czech Republic

18:00 Welcome party

Tuesday, 30 May

9:00-12:00

Guijarro, J.A. (ES): Homogenization of a dense thermo-pluviometric monthly database in the Balearic Islands using the free contributed R package "climatol"

Staudt, M., Esteban-Parra, M.J. and Castro-Díez, Y. (ES): Obtaining a homogenized dataset of monthly Spanish maximum and minimum temperatures

González-Hidalgo, J.C., De Luis Arrirraga, M., Stepanek, P., Lanjeri, S. (ES, CZ): Quality Control of monthly precipitation series from Mediterranean areas of Spain

Toreti, A. and Desiato, F. (IT): Homogenization and validity controls for temperature trend estimates over Italy

Kejna, M. (PL): Homogenisation of air temperature series from Antarctic

Petrović, P. (SCG): Detection Of Inhomogeneities In Wind Direction And Speed Data

Lunch break

14:00-17:00

Kveton, V. and Zak, M. (CZ): Urban effects on the temperature time series of Prague

Aguilar, E., Brunet, M., Saladié, O., Sigró, J. (ES): Homogenization of the Spanish Daily Temperature Series. A step forward.

Štěpánek, P., Řezníčková, L., Brázdil, R. and Pilařová, Z. (CZ): Homogenization of daily air pressure and temperature series of Brno in the period 1848–2005

Mestre, O., Prieur, C. and Caussinus, H. (FR): Daily temperature homogenisation based on non-parametric kernel regression

Szentimrey, T. (HU): Development of MASH homogenization procedure for daily data

Wednesday, 31 May

Excursion to Lake Balaton

Meeting point: HMS, 1, Kitaibel P. street at **8:00**

Thursday, 01 June

9:00-12:00

Domonkos, P. (HU): Testing of homogenisation methods: purposes, tools and problems of implementation

Venema, V., Rust, H., Bachner, S., Kapala, A. and Simmer, C. (DE): Generators for surrogate historical daily data records withknown statistical properties for testing algorithms

Wan, H., Wang, X.,L. and Swail, V.R. (CA): A Quality Assurance System for Canadian Pressure Data

Niedźwiedź, T. (PL): Selected problems regarding homogenization and quality control of daily and monthly pressure data

Boroneant, C., Baciu, M. and Orzan, A. (RO): On the statistical parameters calculated for the essential climatological variables during 2-years of parallel observations with automatic and classical stations in Romania

Cheval, S., Baciu, M., Copaciu, V., Breza, T. and Pescaru, V. (RO): Intercomparison between the hourly meteorological parameters provided by the automatic and classical stations in Romania

Lunch break

14:00-16:30

Vizi, Zs. and Przybylak, R. (PL): Estimation of the accuracy of methods used for the calculation of mean and extreme daily air temperature values in the American Arctic in the 19TH century

Stepanek, P. and Mihulová, K. (CZ,SK): Homogenization of air temperature and relative humidity monthly means for individual observation hours in the area of the Czech and Slovak Republik

Van Hauteghem, H. (BE): Quality Control Framework at the Royal Meteorological Institute of Belgium

Chen, Y. and Churkina, G. (DE): A comparison of climate variables between various data source as the climate forcing to ecosystem modelling

Heino, R. (FI): CCl perspectives on climate data

18:00 Seminar banquet

Friday, 02 June (only morning session)

9:00-13:00

Discussion:

- COST action plan
- Further cooperation and plans
- recommendations

Additionally:

PC experience, Presentation of softwares

LIST OF PARTICIPANTS

BELGIUM

HANS VAN HAUTEGHEM
Royal Meteorological Institute of Belgium
hans@oma.be

BULGARIA

TANIA MARINOVA
National Institute of Meteorology and Hydrology
Tania.Marinova@meteo.bg

CANADA

XIAOLAN (LEONA) WANG
Climate Research Division,
ASTD, STB, Environment Canada
Xiaolan.Wang@ec.gc.ca

EVA MEKIS
Meteorological Sevice of Canada
Eva.Mekis@ec.gc.ca

CROATIA

ANA SANTIC
Meteorological and Hydrological Service of Croatia
santic@cirus.dhz.hr

KSENIJA ZANINOVIC
Meteorological and Hydrological Service of Croatia
zaninovic@cirus.dhz.hr

CZECH REPUBLIC

VIT KVETON
Czech Hydrometeorological Institute
vit.kveton@chmi.cz

MICHAL ZAK
Czech Hydrometeorological Institute
michal.zak@chmi.cz

PETR ŠTĚPÁNEK
Czech Hydrometeorological Institute, regional office Brno
petr.stepanek@chmi.cz

LADISLAVA ŘEZNÍČKOVÁ
Institute of Geography, Masaryk University
reznickova@mail.muni.cz

FINLAND

RAINO HEINO
Finnish Meteorological Institute
raino.heino@fmi.fi

FRANCE

OLIVIER MESTRE
Météo-France
olivier.mestre@meteo.fr

SYLVIE JOURDAIN
Météo-France
Direction de la Climatologie
sylvie.jourdain@meteo.fr

GERMANY

HERMANN MÄCHEL
German Weather Service
hermann.maechel@dwd.de

YOUMIN CHEN
Max-Planck-Institute for Biogeochemistry
ychen@bgc-jena.mpg.de

VICTOR VENEMA
Meteorologisches Institut, Universität Bonn, Germany
Victor.Venema@uni-bonn.de

GREECE

MAGDALINI TZANAKOU
Hellenic National Meteorological Service
magdalini.tzanakou@gmail.com, klimat1@hnms.gr

HUNGARY

BLANKA BARTÓK
University of Debrecen
bartokblanka@yahoo.com

PETER DOMONKOS
dpeterfree@freemail.hu

ZITA BIHARI
Hungarian Meteorological Service
bihari.z@met.hu

BÁLINT BIRSZKI
Hungarian Meteorological Service
birszki.b@met.hu

MÓNIKA LAKATOS
Hungarian Meteorological Service
lakatos.m@met.hu

SÁNDOR SZALAI
Hungarian Meteorological Service
szalai.s@met.hu

TAMÁS SZENTIMREY
Hungarian Meteorological Service
szentimrey.t@met.hu

IRELAND

SÉAMUS WALSH
Irish Meteorological Service
seamus.walsh@met.ie

ITALY

ANDREA TORETI
Agenzia per la protezione dell'ambiente e per i servizi tecnici (APAT)
andrea.toreti@apat.it

FRANCO DESIATO
Agenzia per la protezione dell'ambiente e per i servizi tecnici (APAT)
franco.desiato@apat.it

STEFANIA VERGARI
Italian Air Force Met Service
vergari@meteoam.it

LATVIA

LJUBA PIROZENOKA
Latvian Environment, Geology and Meteorology Agency
ljuba.pirozenoka@meteo.lv

POLAND

JERZY CZECHOWICZ
Data Base Center
Institute of Meteorology and Water Management
jerzy.czechowicz@imgw.pl

MAREK KEJNA
Department of Climatology, Nicholas Copernicus University
makej@geo.uni.torun.pl

ZSUZSANNA VIZI
Department of Climatology, Nicholas Copernicus University
vizi@uni.torun.pl

TADEUSZ NIEDZWIEDZ
University of Silesia, Faculty of Earth Science, Department of Climatology
niedzwie@ultra.cto.us.edu.pl

ROMANIA

SORIN CHEVAL
National Meteorological Administration
sorincheval@fulbrightweb.org

CONSTANTA BORONEANT
National Meteorological Administration
boroneant@meteo.inmh.ro

SERBIA-MONTENEGRO

PETROVIĆ PREDRAG
Republic Hydrometeorological Service of Serbia
p.petrovic@hidmet.sr.gov.yu

SLOVAK REPUBLIC

KATARÍNA MIKULOVÁ
Slovak Hydrometeorological Institute
katarina.mikulova@shmu.sk

SLOVENIA

GREGOR VERTACNIK
Environmental Agency of the Republic of Slovenia
gregor.vertacnik@gov.si

SOUTH AFRICA

ERNEST MBAMBALALA
ernest@weathersa.co.za

SPAIN

MATTHIAS STAUDT
Atmosphere Physics Group, Applied Physics Department, University of Granada
mstaudt@ugr.es,
traduc@matist.jazztel.es

ENRIC AGUILAR
Climate Change Research Group, Geography Unit, Universitat Rovira i Virgili, Tarragona
enric.aguilar@urv.net

JOSÉ CARLOS GONZÁLEZ HIDALGO
Department of Geography, Zaragoza University
jcgh@unizar.es; mdla@unizar.es

SIHAM LANJERI
Department of Geography, Zaragoza University
siham.lanjeri@unizar.es

JOSÉ ANTONIO GUIJJARO
Instituto Nacional de Meteorología
jaguijarro@inm.es

SYRIA

MUHAMMAD IBRAHIM
Syrian Meteorological Department

REPORTS PUBLISHED IN THE

WORLD CLIMATE DATA PROGRAMME (WCDP)/

WORLD CLIMATE DATA AND MONITORING PROGRAMME (WCDMP) SERIES

WCDP-1 WMO REGION III/IV TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT AND USER SERVICES, Barbados, 22-26 September 1986 and Panama, 29 September 3 October 1986 (available in English and Spanish) - (WMO-TD No. 227)

WCDP-2 REPORT OF THE INTERNATIONAL PLANNING MEETING ON CLIMATE SYSTEM MONITORING, Washington DC, USA, 14-18 December 1987 - (WMO-TD No. 246)

WCDP-3 GUIDELINES ON THE QUALITY CONTROL OF DATA FROM THE WORLD RADIOMETRIC NETWORK, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 258)

WCDP-4 INPUT FORMAT GUIDELINES FOR WORLD RADIOMETRIC NETWORK DATA, Leningrad 1987 (prepared by the World Radiation Data Centre, Voeikov Main Geophysical Observatory) - (WMO-TD No. 253. p. 35)

WCDP-5 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS, 1989 edition (WMO-TD No. 293)

WCDP-6 CLICOM PROJECT (Climate Data Management System), April 1989 (updated issue of WCP-l 1 9) - (WMO-TD No. 299)

WCDP-7 STATISTICS ON REGIONAL NETWORKS OF CLIMATOLOGICAL STATIONS (based on the INFOCLIMA World Inventory). VOLUME II: WMO REGION I - AFRICA (WMO-TD No. 305)

WCDP-8 INFOCLIMA CATALOGUE OF CLIMATE SYSTEM DATA SETS - HYDROLOGICAL DATA EXTRACT, April 1989 - (WMO-TD No. 343)

WCDP-9 REPORT OF MEETING OF CLICOM EXPERTS, Paris, 11-15 September 1989 (available in English and French) - (WMO-TD No. 342)

WCDP-10 CALCULATION OF MONTHLY AND ANNUAL 30-YEAR STANDARD NORMALS, March 1989 (prepared by a meeting of experts, Washington DC, USA) - (WMO-TD No. 341)

WCDP-11 REPORT OF THE EXPERT GROUP ON GLOBAL BASELINE DATASETS, Asheville, USA, 22-26 January 1990 - (WMO-TD No. 359)

WCDP-12 REPORT OF THE MEETING ON HISTORICAL ARCHIVAL SURVEY FOR CLIMATE HISTORY, Paris, 21-22 February 1990 - (WMO-TD No. 372)

WCDP-13 REPORT OF THE MEETING OF EXPERTS ON CLIMATE CHANGE DETECTION PROJECT, Niagara-on-the-Lake, Canada, 26-30 November 1990 - (WMO-TD No. 418)

Note: ***Following the change of the name of the World Climate Data Programme (WCDP) to World Climate Data and Monitoring Programme (WCDMP) by the Eleventh WMO Congress (May 1991), the subsequent reports in this series will be published as WCDMP reports, the numbering being continued from No. 13 (the last 'WCDP" report).***

WCDMP-14 REPORT OF THE CCl WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 21-25 October 1991

WCDMP-15 REPORT OF THE CCl EXPERTS MEETING ON CLIMATE CODE ADAPTATION, Geneva, 5-6 November 1991 - (WMO-TD No. 468)

WCDMP-16 REPORT OF THE CCl EXPERTS MEETING ON TRACKING AND TRANSMISSION OF CLIMATE SYSTEM MONITORING INFORMATION, Geneva, 7-8 November 1991 - (WMO-TD No. 465)

WCDMP-17 REPORT OF THE FIRST SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 19-20 November 1991 (also appears as WCASP-18) - (WMO-TD No. 475)

WCDMP-18 CCl WORKING GROUP ON CLIMATE DATA, Geneva, 11-15 November 1991 (WMO-TD No. 488)

WCDMP-19 REPORT OF THE SECOND CLICOM EXPERTS MEETING, Washington DC, 18-22 May 1992 - (WMO-TD No. 511)

WCDMP-20 REPORT ON THE INFORMAL PLANNING MEETING ON STATISTICAL PROCEDURES FOR CLIMATE CHANGE DETECTION, Toronto, 25 June, 1992 (WMO-TD No. 498)

WCDMP-21 FINAL REPORT OF THE CCI WORKING GROUP ON CLIMATE DATA AND ITS RAPPORTEURS, November 1992 - (WMO-TD No. 523)

WCDMP-22 REPORT OF THE SECOND SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 16-17 November 1992 (also appears as WCASP-22) - (WMO-TD No. 529)

WCDMP-23 REPORT OF THE EXPERTS MEETING ON REFERENCE CLIMATOLOGICAL STATIONS (RCS) AND NATIONAL CLIMATE DATA CATALOGUES (NCC), Offenbach am Main, Germany, 25-27 August 1992 - (WMO-TD No. 535)

WCDMP-24 REPORT OF THE TENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Geneva, 20-22 September 1995 (also appears as WCASP-34) - (WMO-TD No. 711)

WCDMP-25 REPORT OF THE FIFTH SESSION OF THE ADVISORY COMMITTEE ON CLIMATE APPLICATIONS AND DATA (ACCAD), Geneva, 26 September 1995 (also appears as WCASP-35) - (WMO-TD No. 712)

WCDMP-26 REPORT ON THE STATUS OF THE ARCHIVAL CLIMATE HISTORY SURVEY (ARCHISS) PROJECT, October 1996 (prepared by Mr M. Baker) - (WMO-TD No. 776)

WCDMP-27 SUMMARY REPORT OF THE MEETING OF THE THIRD SESSION OF THE CCl WORKING GROUP ON CLIMATE CHANGE DETECTION, Geneva, 26 February - 1 March 1996 - (WMO-TD No. 818)

WCDMP-28 SUMMARY NOTES AND RECOMMENDATIONS FOR CCI-XII FROM MEETINGS CONVENED TO PREPARE FOR PUBLISHING THE FIFTH AND SIXTH GLOBAL CLIMATE SYSTEM REVIEWS AND FOR A PUBLICATION ON THE CLIMATE OF THE 20TH CENTURY, July 1997 - (WMO-TD No. 830)

WCDMP-29 CLIMATE CHANGE DETECTION REPORT - REPORTS FOR CCI-XlI FROM RAPPORTEURS THAT RELATE TO CLIMATE CHANGE DETECTION, July 1997 (WMO-TD No. 831)

WCDMP-30 SUMMARY NOTES AND RECOMMENDATIONS ASSEMBLED FOR CCI-XlI FROM RECENT ACTIVITIES CONCERNING CLIMATE DATA MANAGEMENT, July 1997 (WMO-TD No. 832)

WCDMP-31 REPORTS FOR CCI-XlI FROM RAPPORTEURS THAT RELATE TO CLIMATE DATA MANAGEMENT, July 1997 - (WMO-TD No. 833)

WCDMP-32 PROGRESS REPORTS TO CCl ON STATISTICAL METHODS, July 1997 (prepared by Mr Christian-Dietrich Schönwiese) (WMO-TD No 834)

WCDMP-33 MEETING OF THE CCl WORKING GROUP ON CLIMATE DATA, Geneva, 30 January - 3 February 1995 - (WMO-TD No. 841)

WCDMP-34 EXPERT MEETING TO REVIEW AND ASSESS THE ORACLE-BASED PROTOTYPE FOR FUTURE CLIMATE DATABASE MANAGEMENT SYSTEM (CDBMS), Toulouse, France, 12-16 May 1997 - (WMO-TD No. 902)

WCDMP-35 REPORT OF THE ELEVENTH SESSION OF THE ADVISORY WORKING GROUP OF THE COMMISSION FOR CLIMATOLOGY, Mauritius, 9-14 February 1998 (also appears as WCASP-47) - (WMO-TD No. 895)

WCDMP-36 REPORT OF THE MEETING OF THE CCl TASK TEAM ON CLIMATE ASPECTS OF RESOLUTION 40, Geneva, Switzerland, 10-1 1 June 1998 - (WMO-TD No. 925)

WCDMP-37 REPORT OF THE MEETING OF THE JOINT CCl/CLIVAR TASK GROUP ON CLIMATE INDICES, Bracknell, UK, 2-4 September 1998 - (WMO-TD No. 930)

WCDMP-38 REPORT OF THE MEETING OF THE WMO COMMISSION FOR CLIMATOLOGY (CCl) TASK GROUP ON A FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEM (CDMS), Ostrava, Czech Republic, 10-13 November 1998 and FOLLOW-UP WORKSHOP TO THE WMO CCl TASK GROUP MEETING ON A FUTURE WMO CDMS, Toulouse, France, 30 March-1 April 1999 - (WMO-TD No. 932)

WCDMP-39 REPORT OF THE MEETING OF THE CCl WORKING GROUP ON CLIMATE DATA, Geneva, Switzerland, 30 November-4 December 1998 - (WMO-TD No. 970)

WCDMP-40 REPORT OF THE MEETING ON CLIMATE STATISTICS, PRODUCT DEVELOPMENT AND DATA EXCHANGE FOCUSING ON CLICOM 3.1, Geneva, 25-29 January 1999 - (WMO-TD No. 971)

WCDMP-41 PROCEEDINGS OF THE SECOND SEMINAR FOR HOMOGENIZATION OF SURFACE CLIMATOLOGICAL DATA, Budapest, Hungary, 9-13 November 1998 (WMO-TD No. 962)

WCDMP-42 REPORT OF THE MEETING OF EXPERTS ON THE CLIMATE OF THE 20TH CENTURY, Geneva, 26-30 April 1999 - (WMO-TD No. 972)

WCDMP-43 REPORT OF THE TRAINING SEMINAR ON CLIMATE DATA MANAGEMENT FOCUSING ON CLICOM/CLIPS DEVELOPMENT AND EVALUATION, Niamey, Niger, 03 May-10 July 1999, (WMO-TD No. 973)

WCDMP-44 REPRESENTATIVENESS, DATA GAPS AND UNCERTAINTIES IN CLIMATE OBSERVATIONS, Invited Scientific Lecture given by Chris Folland to the WMO Thirteenth Congress, Geneva, 21 May 1999 - (WMO-TD No. 977)

WCDMP-45 WORLD CLIMATE PROGRAMME - WATER, DETECTING TREND AND OTHER CHANGES IN HYDROLOGICAL DATA, Zbigniew W. Kundzewicz and Alice Robson (Editors) - (WMO-TD No. 1013)

WCDMP-46 MEETING OF THE WMO CCl TASK GROUP ON FUTURE WMO CLIMATE DATABASE MANAGEMENT SYSTEMS (CDMSs), Geneva, 3-5 May 2000 (WMO-TD No. 1025)

WCDMP-47 REPORT ON THE ACTIVITIES OF THE WORKING GROUP ON CLIMATE CHANGE DETECTION AND RELATED RAPPORTEURS, 1998-2001 (May 2001, updated from March 2001) (WMO-TD No. 1071)

WCDMP-48 REPORT OF THE FIRST SESSION OF THE MANAGEMENT GROUP OF THE COMMISSION FOR CLIMATOLOGY (Berlin, Germany, 5-8 March 2002) (also appears as WCASP-55) (WMO-TD No. 1110)

WCDMP-49 1. REPORT ON THE CLICOM-DARE WORKSHOP (San José, Costa Rica, 17-28 July 2000); 2. REPORT OF THE INTERNATIONAL DATA RESCUE MEETING (Geneva, 11-13 September 2001) (WMO-TD No. 1128)

WCMDP-50 REPORT OF THE CLIMATE DATABASE MANAGEMENT SYSTEMS EVALUATION WORKSHOP (Geneva, 11-13 September 2001) (WMO-TD No. 1130)

WCDMP-51 SUMMARY REPORT OF THE EXPERT MEETING FOR THE PREPARATION OF THE SEVENTH GLOBAL CLIMATE SYSTEM REVIEW (7GCSR) (Geneva, 16-19 September 2002) (WMO-TD No. 1131)

WCDMP-52 GUIDELINES ON CLIMATE OBSERVATION NETWORKS AND SYSTEMS (WMO-TD No. 1185)

WCDMP-53 GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION (WMO-TD No. 1186)

WCDMP-54 REPORT OF THE CCl/CLIVAR EXPERT TEAM ON CLIMATE CHANGE DETECTION, MONITORING AND INDICES (ETCCDMI) (Norwich, UK, 24-26 November 2003) (WMO-TD No. 1205)

WCDMP-55 GUIDELINES ON CLIMATE DATA RESCUE (WMO-TD No. 1210)

WCDMP-56 FOURTH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES (Budapest, Hungary, 6-10 October 2003) (WMO-TD No. 1236)

WCDMP-57 REPORT OF THE RA V DATA MANAGEMENT WORKSHOP (Melbourne, Australia, 28 November-3 December 2004) (WMO-TD No. 1263)

WCDMP-58 GUIDELINES ON CLIMATE WATCHES (WMO-TD No. 1269)

WCDMP-59 REPORT OF THE MEETING OF THE RA I WORKING GROUP ON CLIMATE MATTERS (Dakar, Senegal, 22 – 24 February 2006) (WMO-TD No. 1351)

WCDMP-60 GUIDELINES ON CLIMATE DATA MANAGEMENT (WMO-TD No. 1376)

WCDMP-61 THE ROLE OF CLIMATOLOGICAL NORMALS IN A CHANGING CLIMATE (WMO-TD No. 1377)

WCDMP-62 GUIDELINES FOR MANAGING CHANGES IN CLIMATE OBSERVATION PROGRAMMES (WMO-TD No. 1378)

WCDMP-63 RA VI TRAINING SEMINAR ON CAPACITY BUILDING IN CLIMATE-RELATED MATTERS (Yerevan, Armenia, 2 – 5 October 2006) (WMO-TD No. 1386)

WCDMP-64 JOINT CCL/CLIVAR/JCOMM EXPERT TEAM ON CLIMATE CHANGE DETECTION AND INDICES (Niagara-on-the-Lake, Canada, 14 - 16 November 2006) (WMO-TD No. 1402)

WCDMP-65 EXPERT TEAM ON OBSERVING REQUIREMENTS AND STANDARDS FOR CLIMATE (Geneva, 28 - 30 March 2007) (WMO-TD No. 1403)

WCDMP-66 A CASE-STUDY/GUIDANCE ON THE DEVELOPMENT OF LONG-TERM DAILY ADJUSTED TEMPERATURE DATASETS (WMO-TD-1425)

WCDMP-67 PROCEEDINGS OF THE INTERNATIONAL WORKSHOP ON RESCUE AND DIGITIZATION OF CLIMATE RECORDS IN THE MEDITERRANEAN BASIN (Tarragona, Spain, 28-30 November 2007) (WMO-TD-1432)

WCDMP-68 CLIMATE DATA MANAGEMENT GUIDELINES

WCDMP-69 REPORT OF THE MEETING OF THE CCL EXPERT TEAM ON THE RESCUE, PRESERVATION AND DIGITIZATION OF CLIMATE RECORDS (Bamako, Mali, 13-15 May 2008) (WMO-TD-1480)

WCDMP-70 GUIDELINES FOR PLANT PHENOLOGICAL OBSERVATIONS (WMO-TD No. 1484)

WCDMP-71 PROCEEDINGS OF THE FIFTH SEMINAR FOR HOMOGENIZATION AND QUALITY CONTROL IN CLIMATOLOGICAL DATABASES (Budapest, Hungary, 29 May-2 June 2006) (WMO-TD-1493)

For more information, please contact:

World Meteorological Organization

Observing and Information Systems Department

Tel.: +41 (0) 22 730 82 68 – Fax: +41 (0) 22 730 80 21

E-mail: wcdmp@wmo.int

7 bis, avenue de la Paix – P.O. Box 2300 – CH 1211 Geneva 2 – Switzerland

www.wmo.int